

Direction des bibliothèques

AVIS

Ce document a été numérisé par la Division de la gestion des documents et des archives de l'Université de Montréal.

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

This document was digitized by the Records Management & Archives Division of Université de Montréal.

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

**Détection, caractérisation et visualisation des structures
transitoires de protéines par sondage au tryptophane**

par

Alexis Vallée-Bélisle

Département de Biochimie

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Docteur
en Biochimie

Janvier, 2008

© Alexis Vallée-Bélisle, 2008



Université de Montréal

Faculté de Médecine

Cette thèse intitulée :

Détection, caractérisation et visualisation des structures transitoires de protéines par
sondage au tryptophane

présentée par :

Alexis Vallée-Bélisle

a été évaluée par un jury composé des personnes suivantes :

Dr Serguei Chteinberg, président-rapporteur

Dr Stephen W. Michnick, directeur de recherche

Dr Jeffrey W. Keillor, membre du jury

Dr Heinrich Roder, examinateur externe

Dr Normand Brisson, représentant du doyen de la faculté

Résumé

Au cours des années 1980, le développement des techniques de clonage et de mutagenèse dirigée (génie génétique) a véritablement propulsé la biochimie dans une ère des plus excitantes en donnant naissance à une nouvelle discipline : le génie des protéines. Ainsi, depuis maintenant plus de deux décennies, les fonctions d'un nombre croissant de protéines naturelles sont déterminées et cette connaissance est mise à profit dans des domaines aussi divers que la médecine, les biotechnologies ainsi les nanotechnologies. Toutefois, malgré les énormes avancées réalisées, nos connaissances des mécanismes de repliement ou de fonctionnement des protéines s'avèrent encore trop limitées pour être en mesure de prédire la structure de protéines à partir de leur séquence d'acide aminé ou de permettre leur design rationnel.

Pour parvenir à comprendre les mécanismes de repliement ou de fonctionnement d'une protéine, il importe de caractériser les différents états intermédiaires empruntés par celle-ci au niveau de la participation énergétique de chacun de ses acides aminés. Seul ce niveau de compréhension rendra possible la mise au point d'algorithmes théoriques suffisamment simples et réalistes pour modéliser efficacement le comportement des protéines à partir de leur séquence d'acide aminés. Toutefois, l'étude expérimentale des mécanismes de repliement et de fonctionnement des protéines comporte actuellement plusieurs limitations allant de la détection à la caractérisation des différentes structures transitoires empruntées par la protéine. En effet, les états transitoires intermédiaires des protéines possèdent généralement des demi-vies beaucoup trop petites pour permettre la détermination de leur structure via les techniques à fortes résolutions structurales (e.g. RMN et cristallographie aux rayons X). Les méthodes suffisamment sensibles pour permettre la détection de changements structuraux rapides (fluorescence, dichroïsme circulaire...) ont, pour leur part, une résolution structurale relativement faible et peu détaillée. À l'heure actuelle, il importe donc de mettre au point de nouvelles techniques ou

stratégies qui permettront de détecter et caractériser de manière efficace les intermédiaires transitoires de protéines au niveau de l'implication de chaque acide aminé.

Dans le présent ouvrage, je détaille les bases et principes d'une nouvelle méthode d'analyse, permettant de contourner certaines limitations actuelles concernant la détection et la caractérisation des structures transitoires de protéines de courte demi-vie : la méthode par sondage au tryptophane. Cette approche combine deux outils couramment utilisés par les ingénieurs de protéines: la spectroscopie de fluorescence ainsi que la mutagenèse dirigée. Dans cette thèse, cette approche est appliquée à l'étude du mécanisme de repliement de la protéine ubiquitine.^a Cette stratégie a permis de détecter, caractériser et visualiser un nouvel état intermédiaire dans le mécanisme de repliement d'ubiquitine là où toutes les autres approches expérimentales classiques ne furent pas en mesure de détecter quoi que ce soit. Notre approche expérimentale permet donc de mieux caractériser les différents mécanismes de protéines et de générer de nombreux paramètres expérimentaux sur les structures transitoires de protéines ce qui contribuera grandement à tester et valider les simulations théoriques de protéines.

Mots-clés : Protéines, mécanisme de repliement, ubiquitine, structure d'intermédiaires transitoires, spectroscopie de fluorescence, mutagenèse dirigée, sondage par tryptophane, analyse par valeur- ω , biotechnologies.

^a Elle aurait également pu servir à caractériser les intermédiaires qui s'accumulent lors du mécanisme de fonctionnement d'une protéine.

Abstract

Proteins can exist in a variety of stable native, or transient intermediate states during folding, in response to post-translational modifications or binding to other molecules. The identification and structural characterization of transient intermediate states is a major challenge and high resolution structural methods (e.g. NMR or x-rays crystallography) do not have the temporal resolution to allow for their detection. However, fluorescence spectroscopy, using intrinsic fluorescence of tryptophan residues, can be used to detect the existence of transient intermediates for most proteins with a temporal resolution down to hundreds of nanoseconds, but with no structural resolution. Here we demonstrate that solvent exposed amino acids on a protein can be substituted for tryptophan residues with minimal perturbation of a protein stability or folding landscape. These numerous fluorescent probes can then be used to characterize the kinetics and energetics of structural transitions as well as characterizing the structures of the transient-states by comparing the fluorescence signal of a tryptophan in an unknown state to known reference states of the protein (i.e. unfolded and folded states) at many different locations. Herein, we provide the spatial characterization of a late intermediate state detected during the folding of the small protein ubiquitin by analyzing the refolding kinetic of 27 different solvent-exposed tryptophan mutants. This approach may further open the door to accurate structural determination of transient states by providing numerous structural constraints derived from the fluorescence signal of the many introduced tryptophans.

Keywords : Proteins, folding mechanism, ubiquitin , transient intermediates state, structure, fluorescence spectroscopy, mutagenesis, tryptophan scanning, ω -value analysis, biotechnology.

Table des matières

Liste des tableaux	ix
Liste des figures.....	x
Remerciements	xv
Abbreviations	xviii
Avant-propos.....	xxi
Chapitre 1. Introduction générale.....	1
1.1. Le mécanisme de repliement des protéines	8
1.1.1. Les états natif et déplié d'une protéine.....	11
1.1.2. L'état de transition d'une réaction de repliement de type « deux états »	21
1.1.3. Les intermédiaires de repliement	26
1.2. Détection, caractérisation, et visualisation des structures transitoires de protéines par sondage au tryptophane.....	33
1.2.1. La stratégie par sondage au tryptophane	34
1.3. Protéines à l'étude, hypothèses et objectifs.....	39
1.4. Références	43
Chapitre 2. Article #1	51
2.1. Abstract	52
2.2. Introduction	53

2.3. Materials and Methods	55
2.4. Results	61
2.5. Discussion.....	68
2.6. Tables	74
2.7. Figures.....	76
2.8. References	87
 Chapitre 3. Article #2	 93
3.1. Abstract	94
3.2. Introduction	96
3.3. Results	99
3.4. Discussion.....	107
3.6. Materials and methods	114
3.7. Tables	117
3.8. Figures.....	120
3.9. References	129
 Chapitre 4. Article #3	 136
4.1. Abstract	138
4.2. Introduction	139
4.3. Results	140
4.4. Discussion.....	144
4.5. Table.....	147
4.6. Figures.....	149
4.7. References	158

Chapitre 5. Discussion	162
5.1. Mécanisme de repliement d'ubiquitine	164
5.1.1. Modèle du repliement d'ubiquitine avant les travaux de cette thèse	165
5.1.2. Nouveau modèle de repliement d'ubiquitine.....	168
5.1.3. Expériences à réaliser dans un futur rapproché	176
5.1.4. Validation des simulations théoriques à partir de modèles simplifiés	180
5.2. Avantages, limitations, et perspectives de la méthode de sondage des structures protéiques par tryptophane	181
5.2.1. Détection et caractérisation des structures transitoires de protéines	181
5.2.1. Visualisation des structures transitoires de protéines.....	183
5.2.3 Autres applications envisagées de l'approche par sondage global au tryptophane	187
5.3. Références	189
 Annexe	 I

Liste des tableaux

Tableau 1-1. Techniques expérimentales permettant d'obtenir de l'information structurale sur les états natifs et non-natifs des protéines.....	31
Table 2-1. Folding Thermodynamic and Kinetic Parameters of mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD.....	74
Table 2-2. Folding Thermodynamic and Kinetic Parameters of <i>raf</i> RBD at 8 °C with 0.4 M Na ₂ SO ₄ and of <i>raf</i> RBD with Different Central Core Residues at Position 82.....	75
Table 3-1. Stabilities and <i>m</i> -values of Ub ^{M1W} , Ub ^{A28W} , Ub ^{F45W} , Ub ^{S57W} , and Ub ^{T66W} obtained using equilibrium, folding, or unfolding experiments or kinetic unfolding experiments.....	117
Table 3-2. Rate constants and amplitudes of the transitions detected during the refolding of Ub ^{wt} , Ub ^{M1W} , Ub ^{A28W} , Ub ^{F45W} , Ub ^{S57W} , and Ub ^{T66W} in 0.45 M GdnHCl using Trp or ANS as fluorescent probes.	118
Table 3-3. Folding and unfolding parameters obtained from a three-state on-pathway global fit analysis of <i>k</i> _{f1} , <i>k</i> _{f2} , <i>k</i> _u and <i>F</i> ₁ for Ub ^{M1W} , Ub ^{S57W} , and Ub ^{F45W-T66W}	119
Table 4-1. Thermodynamic and kinetic parameters obtained for all Trp mutants.	147

Liste des figures

Figure A. Extrait de mon cahier « Biologie Moléculaire #6 » du 5 avril 2002.	xx
Figure 1-1. Représentation schématique de diverses protéines.....	3
Figure 1-2. Les différents états d'une protéine et sa surface énergétique.....	5
Figure 1-3. Comment étudier le mécanisme de repliement des protéines en six étapes.....	10
Figure 1-4. Coopérativité de la transition de repliement des petites protéines naturelles. ..	12
Figure 1-5. Forces et motifs permettant de stabiliser l'état natif des protéines et propension des acides aminés à se retrouver à certaines positions.....	15
Figure 1-6. Ensembles de structures représentant les états natif et déplié de protéines.	20
Figure 1-7. État de transition d'un mécanisme de repliement à « 2 états ».	25
Figure 1-8. Les états transitoires détectés lors du repliement des protéines.	28
Figure 1-9. L'acide aminé tryptophane : structure et caractéristique de fluorescence.	36
Figure 1-10. Caractérisation et visualisation des intermédiaires transitoires de repliement.	38
Figure 1-11. Ubiquitine : structure native, historique des principales découvertes et trace de repliement.	42
Figure 2-1. Tertiary and secondary ribbon structures of mUbi and <i>raf</i> RBD and their primary structure alignment.....	76
Figure 2-2. Fluorescence emission spectra of yUbi ^{F45W} and <i>raf</i> RBD.	77

Figure 2-3. Fraction folded vs denaturant and free energy of unfolding of mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD at 25 °C.....	78
Figure 2-4. Stopped-flow kinetic refolding traces of mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD at different denaturant concentrations.....	79
Figure 2-5. Extrapolated initial and final fluorescence intensity observed during folding and unfolding kinetic experiments with mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD.....	80
Figure 2-6. Kinetic relaxation rates of folding and unfolding transitions of mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD.....	81
Figure 2-7. Effect of double jump (2ble-jump, rapid unfolding and refolding) on the refolding heterogeneity of His-mUbi ^{F45W} and His- <i>raf</i> RBD.....	82
Figure 2-8. Comparative chevron curve plots of mUbi ^{F45W} , yUbi ^{F45W} , and <i>raf</i> RBD at 25 C.	83
Figure 2-9. Equilibrium stability curves and chevron curves of <i>raf</i> RBD at 8 °C and with 0.4 M Na ₂ SO ₄	84
Figure 2-10. Equilibrium stability curves and chevron curves of <i>raf</i> RBD mutants L82I, L82V, and L82A.	85
Figure 2-11. Comparison of differences in the free energy barrier of folding and unfolding obtained when various perturbations are used to affect folding-unfolding pathways for both <i>raf</i> RBD and mUbi.	86
Figure 3-1. Ub refolding probed by Trp 45 fluorescence and distinguishing between sequential and parallel folding intermediates using multiple tryptophan probes.	120
Figure 3-2. Equilibrium characterization of the different single Trp mutants.	121

Figure 3-3. Unfolding kinetic characterization of the different single Trp mutants.....	122
Figure 3-4. Folding kinetic traces of the different Trp mutants monitored using ANS fluorescence.	123
Figure 3-5. Folding kinetic characterization of the different single Trp mutants monitored using Trp fluorescence.	124
Figure 3-6. Folding kinetic characterization of the double Trp mutant Ub ^{F45W-T66W} monitored using Trp fluorescence.	125
Figure 3-7. Characterization of Ub late I-state using folding/unfolding double-jump experiments on Ub ^{M1W}	126
Figure 3-8. Free energy diagram for Ub folding/unfolding.	127
Figure 3-9. Refolding traces and rate constants of the two transitions detected between 3.2 and 300 ms when refolding in different GdnHCl concentration, at different protein concentrations and at pH 7.0 and 5.0.	135
Figure 4-1. Characterizing protein transient-states using Trp scanning mutagenesis.	149
Figure 4-2. Transient intermediate states detected during ubiquitin (Ub) refolding.	150
Figure 4-3. Equilibrium and kinetic characterization of the different Ub Trp mutants.	
Figure 4-4. Fluorescence-based structural representation of the Ub late folding intermediate.	152
Supplementary Figure 4-1. Equilibrium curves, folding traces at 0.45 M GdnHCl, and rate of the apparent “two-state” folding transition detected above 2 M GdnHCl for all Trp mutants.	153

Supplementary Figure 4-2. Global fitting analysis of the kinetic and thermodynamic parameters obtained for all Trp mutants.	155
Figure 5-1. Schéma illustrant l'état de nos connaissances sur le mécanisme de repliement de la protéine ubiquitine avant les travaux de cette thèse, et suivant les travaux de cette thèse.....	166
Figure 5-2. Les acides aminés faisant varier la vitesse de T_2 par plus de 2x lorsque mutés en tryptophane.....	170
Figure 5-3. Modèle hypothétique de la structure de l'intermédiaire hâtif et de son réarrangement en l'état natif.....	174
Figure 5-4. Position des tryptophanes présentant une fluorescence non-native dans l'état intermédiaire.....	177

À la vie!
(le plus passionnant
des sujets de recherche)

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse le Professeur Stephen Michnick qui m'a transmis sa passion pour la biochimie et les biotechnologies. Je le remercie pour sa patience, son enthousiasme, son ouverture d'esprit et surtout, pour m'avoir appris à penser « grand »! Mon passage dans son laboratoire hautement multi disciplinaire et explosif aura eu un impact déterminant sur ma vie! Je le remercie de m'avoir offert la chance de participer à des congrès scientifiques de haut niveau et de m'avoir assuré un soutien financier tout au long de mes études me permettant ainsi de mener à bien la complétion de mes travaux de doctorat. Cette aide fut d'autant plus précieuse sachant à quel point le financement des recherches fondamentales sur le mécanisme de repliement des protéines est pratiquement inexistant dans notre belle province.

Je remercie le Professeur Tobin Sosnick de l'Université de Chicago pour m'avoir guidé dans l'analyse de mes résultats ainsi que dans la mise au point du nouveau modèle de repliement d'ubiquitine présenté au sein de cette thèse. Les nombreuses heures/nuits/semaines passées à retourner de tous sens, tous côtés, plusieurs de ses arguments relativement aux résultats de cette thèse, ou d'autres articles, auront grandement contribué à simplifier la forme et la structure de mon 2^e article tout en faisant de moi un meilleur chercheur. Je remercie également la Professeure Sheena Radford de l'Université de Leeds pour m'avoir prodigué de précieux conseils concernant la nature de « mon » intermédiaire ainsi que le Professeur Kevin Plaxco de l'Université de Californie à Santa Barbara pour m'avoir permis de collaborer à la mise en place de l'International Foldiomics Consortium. Enfin, merci aux Professeurs Serguei Chtenbergh et Jeffrey Keillor de l'Université de Montréal pour avoir accepté de suivre avec intérêt et rigueur l'avancement de mes travaux via mes comités de thèse et au Professeur Heinrich Roder du Fox Chase Cancer Center à Philadelphie d'avoir eu la gentillesse de bien vouloir agir en tant que membre externe de mon jury de thèse.

Un merci chaleureux à mes collègues et amis de travail qui ont grandement contribué à rendre mon travail en laboratoire aussi stimulant d'un point de vue intellectuel, qu'agréable et divertissant au point de vue humain. À François-Xavier, Jean-François T., Jean-François P., Ingrid, Emily, Po hien, Livia, Eddy, Valérie, Carol, Kirill, Éléna, Mercedes, Stevo, Igor, Ibrahima, Mohan, Vincent, Martin, Galia, Émilie, Melissa, Dimitri, Hugo, Geoffroy, Jérôme, Annie, Joelle, André (et tous les autres...): merci pour votre présence! Merci particulier à François-Xavier de m'avoir accompagné physiquement, moralement, scientifiquement, amicalement et surtout fidèlement tout au long de ce parcours. À Jean-François Turcotte pour m'avoir prêté main forte durant la préparation de mon premier article. À Jean-François Paradis et Po Hien qui ont toujours réussi à nourrir cette ambiance familiale au laboratoire par plein de petites attentions. À mes correctrices préférées Melissa et Emily pour avoir grandement faciliter la publication de mes articles. Au professeur Jeff Keillor et les gens de son laboratoire pour avoir eu la gentillesse de m'ouvrir les portes de leur laboratoire me donnant ainsi accès à cet instrument de mixage (mon principal allié/adversaire!) duquel la grande majorité des résultats de cette thèse sont issus!

Merci au département de biochimie, à sa direction, son corps professoral ainsi que ses employés pour leurs efforts soutenus dans la création d'un sentiment d'appartenance à notre unité de recherche. Merci aux bénévoles de l'association étudiante qui contribuent, plus que tous autres, à cet esprit de corps : Martin Baril, Karine Boulay, Mélissa Léger, Jordan Volpato et Martin Primeau, ce fut un plaisir de travailler en votre compagnie. Un merci tout particulier à Sylvie Beauchemin pour son professionnalisme (toujours là pour sauver notre vie!) et sa bonne humeur à toute épreuve ainsi qu'à Myreille qui m'a rendu de fier service de séquençage.

Enfin, mes remerciements les plus nourris vont à ceux et celles qui par leur amitié et leur amour ont donné un sens à ma passion, contres vents et marées.

Ma famille rapprochée : Lorraine, Serge, Roxane, Laurent, Vincent, Renaud, Louise, Marc, Peter, Marie-Hélène, Victor, Simon (Savanne et Lilou!), parce qu'il y a beaucoup de vous dans cette thèse!

Mes amis musiciens de (n=1) : Mathieu, Philippe et Alexandre; comment aurai-je pu survivre à ces longues années d'études sans nos moments fous de jams/pratiques/enregistrements/shows!

Mes amis de la science, FX, Stéphanie, Dimitri, Émilie, Galia, Ingrid, JF, Po Hien, Edie, Jérôme et tous les autres pour avoir si fidèlement « enduré » toutes mes folies et écarts de conduites au cours de ces années!

Mes amis de longue date, Marie-Isabelle, Laurence, Catherine, Antoine, Murielle, Benoît, Roseline, Hendrick, Myriam, Frédéric, François, Sébastien, Naïma; pour me rappeler régulièrement qu'il y a des choses pas mal plus importantes que la Science dans la vie!

Ma super « gang » de hockey : Pierre, Colin, Paul, Manu, Martin, Denis, Laurence, Pascal, Olivier, Carol et tous les autres; pour m'avoir permis de vous plaquer dans la bande au moins une fois par semaine durant les dix dernières années (Ouf ça faisait donc du bien!).

Mes co-locs : Franck, Hendrick, Catherine, Claude, David et Savane; pour avoir toujours été pour moi une deuxième famille.

Mon oncle Michel et mes ptits cousins Kassandra et Shaman; pour m'avoir permis de me ressourcer dans la nature en votre compagnie.

Et finalement, Richard et Camilla ainsi que les Wetmores pour m'avoir si chaleureusement accueilli chez vous à Shédiac durant la rédaction de ce document à saveur saline! ;)

Abbreviations

(Les noms anglais de certaines méthodes expérimentales particulièrement difficiles à traduire sont placés entre guillemets)

Acides aminés (code 1 lettre/code 3 lettres)

A/Ala:	Alanine
C/Cys:	Cystéine
D/Asp:	Aspartate
E/Glu:	Glutamate
F/Phe:	Phénylalanine
G/Gly:	Glycine
H/His:	Histidine
I/Ile:	Isoleucine
K/Lys:	Lysine
L/Leu:	Leucine
M/Met:	Méthionine
N/Asn:	Asparagine
P/Pro:	Proline
Q/Gln:	Glutamine
R/Arg:	Arginine
S/Ser:	Sérine
T/Thr:	Thréonine

V/Val:	Valine
W/Trp:	Tryptophane
Y/Tyr:	Tyrosine

ANS :	1-anilino-8-naphtalènesulfonate
β_T :	Degré de compaction relative d'un état
[D] :	Concentration de dénaturant
DC :	Dichroïsme circulaire
$\Delta\Delta G_{F-U}$:	Variation de la différence d'énergie libre
$\Delta\Delta G_{U-\ddagger}$:	Variation de la différence d'énergie libre de repliement ($U \rightarrow \ddagger$)
$\Delta\Delta G_{\ddagger-F}$:	Variation de la différence d'énergie libre de dépliement ($F \rightarrow \ddagger$)
ΔG^{H_2O} :	Différence d'énergie libre en absence de dénaturant
ΔG_{F-U} :	Différence d'énergie libre
$\Delta G_{U-\ddagger}$:	Différence d'énergie libre de repliement
$\Delta G_{\ddagger-F}$:	Différence d'énergie libre de dépliement
DER:	« Dynamic ensemble refinement »
DLR :	Domaine liant Ras
É.D. :	État déplié
É.I. :	État intermédiaire
É.N. :	État natif
É.T. ou \ddagger :	État de transition
É.D. \rightarrow É.N.:	Transition de l'état dénaturé vers l'état natif

FRET :	Transfert d'énergie de résonance de Förster
GdnHCl :	Hydrochlorure de Guanidium
K_e :	Constante d'équilibre
k :	Constante de vitesse
IM7 :	Protéine d'immunité liant E colicin type 7
m :	Coefficient de variation en fonction de la [D]
Na_2SO_4 :	Sodium sulfate
Ponts H :	Ponts Hydrogènes
RMN :	Résonance magnétique nucléaire
RMSD :	« Root mean square deviation »
SH3 :	« Src3 homology »
T :	Transition
wt :	Type sauvage

Avant-propos

Ce qui m'apparaît le plus passionnant dans l'étude de la biochimie n'est pas tant cette extraordinaire ingéniosité qu'affiche les innombrables mécanismes moléculaires mis au point par la nature, que toutes les péripéties qu'auront traversées les hommes de science pour arriver à mettre à jour ces merveilles. En ce sens, je crois sincèrement que la lecture de cette thèse auraient été plus excitante si mes résultats y avaient été rapportés de la manière par lesquels ils sont venus à moi au cours de ces longues années d'efforts c'est-à-dire d'une manière tout à fait inattendue, et suivant mille détours! Ainsi, vous auriez pu partager avec moi ces mois d'angoisse passés à replier « ma » première protéine, le DLR de *raf*, avec cet appareil de mixage rapide fraîchement débarqué d'Angleterre : le « stopped-flow ». Mois durant lesquels je ne savais dire si les imperfections dans mes traces de repliement étaient simplement dues à une variation de pression de mon appareil de mixage au environ de 20 millisecondes ou à la présence d'un intermédiaire de repliement transitoire. Vous auriez également vibré avec moi à la lecture du chapitre décrivant la nuit du 5 au 6 avril 2002 au cours de laquelle j'ai réalisé pour la première fois l'ampleur des applications rendu possible en variant simplement la position du tryptophane dans « mon » autre protéine, ubiquitine, pour me permettre de suivre son repliement dans le temps (Figure A).

Au cours de la rédaction de cette thèse, j'ai ainsi réalisé qu'il existait une grande quantité d'angles sous lesquels mes résultats auraient pu être rapportés. Autant de points de vue qui auraient davantage mis en valeur telle hypothèse ou telle conclusion. J'ai finalement décidé de présenter mes résultats en mettant l'emphasis sur ce que je croyais être la facette principale de ma recherche c'est-à-dire, le développement d'une nouvelle stratégie permettant de détecter, caractériser et visualiser les structures transitoires de protéines. Évidemment, cette nouvelle approche devait, au départ, simplement permettre de

mieux comprendre le repliement de « mes » protéines. Mais force est de constater qu'après plusieurs mois d'analyse et de réflexion, cette stratégie m'apparaît comme étant beaucoup plus digne d'intérêt que les résultats en soi!

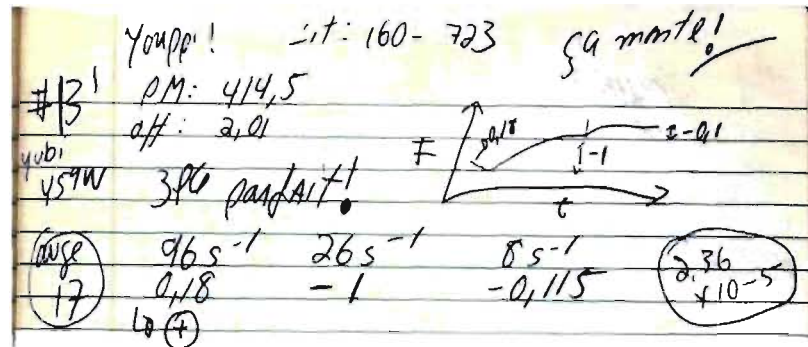


Figure A. Extrait de mon cahier « Biologie Moléculaire #6 » du 5 avril 2002. Commentaires suivant l'analyse de la trace de repliement du premier mutant tryptophane « yUbiY59W » (l'ubiquitine de levure contenant un tryptophane à la position 59). Les mots « Youppi ! » et « 3pe parfait ! » (faisant référence à la présence évidente de trois transitions) reflètent bien l'émotion de l'étudiant au doctorat flairant la bonne affaire!

Cette thèse est divisée en cinq chapitres. Le cœur de celle-ci (chapitres 2 à 4) représente mes trois articles. D'une manière générale, l'article #1 (chapitre 2) rapporte la caractérisation du mécanisme de repliement de trois protéines membres de la superfamille d'ubiquitine à l'aide des méthodes d'analyse classiques du mécanisme de repliement des protéines. Cet article met également en lumière la limitation des différentes approches existantes pour détecter les états intermédiaires transitoires. Le second article (Chapitre 3) propose donc une nouvelle stratégie rendant possible la détection et la caractérisation efficaces de ces états. Enfin, le dernier article présente une nouvelle méthode d'analyse permettant de visualiser la structure de ces intermédiaires transitoires. Comme chacun de mes articles contient déjà une introduction ainsi qu'une discussion spécifique à leur contenu, le premier chapitre de ma thèse servira ainsi simplement d'introduction générale au domaine du repliement des protéines tandis que le cinquième chapitre présentera une

brève discussion résumant les résultats de la thèse tout en discutant des principales applications et perspectives envisagées pour cette nouvelle technologie. Les lecteurs retrouveront finalement en annexe, un manifeste de l'International Foldiomics Consortium auquel j'ai eu le privilège de contribuer au cours de mes études doctorales et dont l'objectif vise à établir des normes standard pour l'étude du mécanisme de repliement des protéines à la communauté scientifique.

En terminant, je me réjouis en pensant que contrairement aux lois fondamentales de thermodynamique concernant les fonctions d'états^b, le chemin parcouru pour parvenir à la complétion de mes travaux doctoraux aura compté mille fois plus à mes yeux que ces quelques centaines de pages blanches tachetées de noir et de quelques couleurs se retrouvant aujourd'hui devant vos yeux... où quelque part, perdu, sur une tablette poussiéreuse au fond d'une bibliothèque !

En espérant que vous prendrez tout de même quelque plaisir à me lire,

Alexis Vallée-Bélisle

^b Les fonctions d'états telles l'énergie libre, l'enthalpie, et l'entropie permettent de déterminer l'état d'un système. L'énergie nécessaire pour passer d'un état A à un état B ne dépend que de l'état initial et de l'état final du système et pas du chemin parcouru durant la transformation.

Chapitre 1. Introduction générale

Depuis des milliards d'années, la nature a mis sur pied une technologie moléculaire incroyablement complexe et ingénieuse rendant possible le développement de la vie sur terre. De toutes les biomolécules, les protéines représentent sans doute la catégorie de molécules déployant la plus grande ingéniosité et diversité de fonctions tant du point de vue structurel que catalytique. Ainsi, à partir de seulement vingt acides aminés différents^c assemblés à la queue leu-leu en des chaînes linéaires de différentes longueurs, et de différentes compositions, les organismes vivants ont réussi à créer une panoplie de machines moléculaires et de matériaux à l'échelle nanoscopique (Figure 1-1). Ces différentes structures ont rendu possible l'établissement de systèmes de communications, de transport, de contrôle de qualité, de transformation de matières premières, de stockage d'énergie, de défense, d'archivage d'information, de construction et de réparation à l'intérieur même des organismes vivants.

Plusieurs raisons outre le simple fait de vouloir comprendre les mécanismes inhérents à la vie, contribuent à accroître de jour en jour l'intérêt que porte la communauté scientifique à l'égard des protéines. Dans un premier temps, les biochimistes mettent en lumière un nombre croissant de maladies ou de pathologies liées à de simples variations d'acides aminés chez des protéines. Par exemple plus de 50% des cancers humains semblent attribuables à des mutations d'acides aminés spécifiques retrouvées au sein de la protéine p53, une protéine permettant la suppression de tumeurs.² Les maladies dégénératives telle la maladie de la vache folle (ou encéphalopathie spongiforme bovine), la maladie de Creutzfeldt-Jacob, la maladie de Parkinson et l'Alzheimer sont toutes causées par des changements structuraux de protéines qui engendrent la formation de polymères ou d'agrégats.³ De nos jours, la compréhension des mécanismes de fonctionnement et de repliement des protéines rend ainsi possible la mise au point de test de dépistage pour de

^c En plus des vingt acides aminés, diverses additions post-traductionnelles permettent également de générer encore plus de variété au niveau de la structure des protéines suivant leur synthèse. Par exemple, ces dernières peuvent être phosphorylées, acétylées, glycosylées, méthylées, hydroxylées, ADP-ribosylées, ubiquitinées...¹

nombreuses maladies et contribue de plus en plus à rendre efficace le design rationnel de nouveaux médicaments permettant d'inhiber ou de compenser l'activité de certaines protéines déficientes.

En plus des retombées qu'elles entraînent sur le secteur de la santé, les recherches effectuées sur les protéines jouent également un rôle de plus en plus déterminant dans l'émergence de nombreuses biotechnologies associées au domaine de l'environnement, de l'agriculture, de l'agro alimentaire ainsi que pour la mise au point de processus industriels innovants. On n'a qu'à penser aux organismes génétiquement modifiés (OGM),⁴⁻⁶ aux biomatériaux à base de protéines,⁷⁻¹⁰ aux enzymes industrielles,¹¹⁻¹³ ainsi qu'au domaine naissant des biosenseurs.¹⁴⁻¹⁶ Ainsi, il devient envisageable de penser que dans un avenir rapproché, l'homme parviendra à maîtriser les technologies ainsi que les règles qui lui permettront de faire le design de nouvelles protéines aux fonctions variées comme on design une voiture ou un ordinateur. Cette connaissance lui fournira du même coup les outils nécessaires lui permettant de soigner les maladies à l'échelle moléculaire comme on répare une machine ou un moteur à l'échelle macroscopique.

Toutefois, la technologie du vivant, issue de millions d'années d'évolution moléculaire, reste encore grandement empreinte de mystère et tarde à nous livrer ses secrets. En effet, la complexité inhérente au design des protéines naturelles repose sur le fait que toute l'information nécessaire à la stabilité de leur structure native, à leur mécanisme de fonctionnement, ainsi qu'à leur mécanisme de repliement, se retrouve entièrement encodé dans la composition de leur chaîne polypeptidique (séquence d'acides aminés).^d Par analogie, c'est comme si une grande majorité des appareils, matériaux, ou technologies mis au point par l'homme étaient composés de seulement vingt éléments de bases qui, une fois joints ensemble à la queue leu-leu, s'assembleraient automatiquement en l'appareil désiré. Encore plus étonnant, les protéines arrivent à accomplir cet exploit

^d Il existe toutefois quelques mécanismes permettant d'assister certaines protéines dans leur mécanisme de repliement à l'intérieur de la cellule.¹⁷

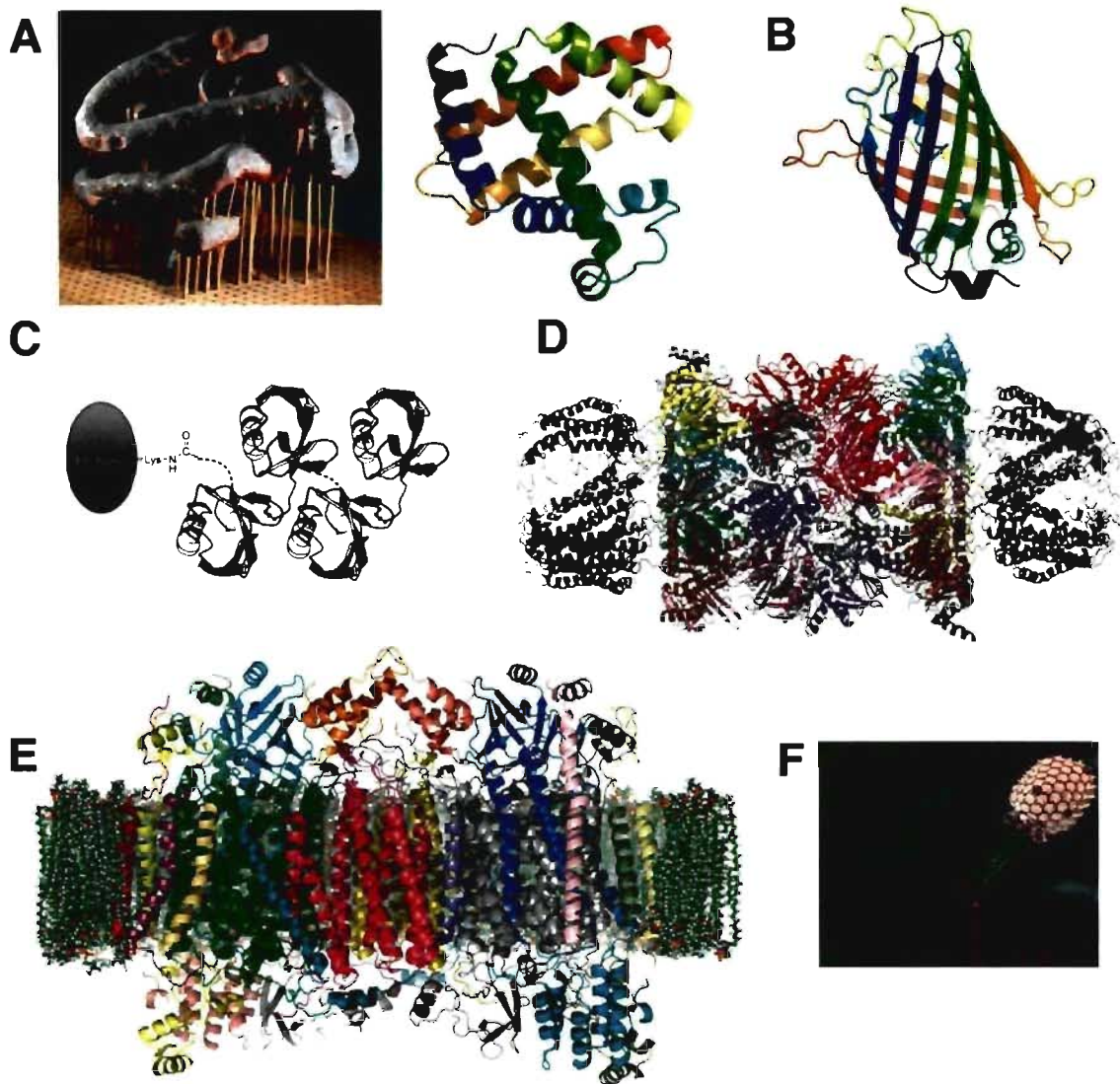


Figure 1-1. Représentation schématique de diverses protéines. A) La myoglobine lie l'oxygène au niveau des muscles. *Gauche* : premier modèle de protéine en plasticine de Kendrew (1957). *Droite* : structure à haute résolution (1mbo).¹⁹ B) Protéine fluorescente verte (PFV) (1ema) permettant aux méduses de simuler un ciel étoilé pour échapper à leur prédateur.²⁰ La PFV a également de nombreuses applications en biotechnologie.²¹ C) Unité de tétraubiquitine (protéine modèle de cette thèse -1ubq) qui lorsque greffée à une protéine cible, achemine cette dernière vers le protéasome (dégradation).²² D) Le protéasome (1fnt) permet de dégrader les protéines mal repliées, dénaturées ou obsolètes de manière ciblée.²³ E) Cytochrome *c* oxidase (2occ), dernière protéine dans la chaîne de transport d'électron qui rend possible la synthèse d'ATP.²⁴ F) Les bactériophages sont des structures protéiques qui rendent possible l'échange de matériel génétique (ADN ou ARN) entre les bactéries.

dans un environnement incroyablement hostile : en présence de centaines voire de milliers d'autres biomolécules comprimées les unes contre les autres.^e Ainsi, pour parvenir à maîtriser un jour l'étendu de la Science et des possibilités offertes par les protéines, les scientifiques doivent donc étudier dans tous ses détails:

- 1) La structure native des protéines naturelles;
- 2) Leur mécanisme de fonctionnement;
- 3) Leur mécanisme de repliement;
- 4) Les états de la protéine qui sont nocifs pour la cellule (par exemple, les états susceptibles de s'agréger ou d'interférer avec le fonctionnement des autres biomolécules);

En bref, le biochimiste doit caractériser la surface énergétique de la protéine, c'est-à-dire caractériser les structures et la dynamique des différents états accessibles à la protéine (Figure 1-2). D'une manière générale, on détermine le mécanisme spécifique d'une protéine (ou de tout autre réaction chimique) en caractérisant les états affichants l'énergie la plus basse et l'énergie la plus haute (état de transition) durant le mécanisme; les états de transition permettant le passage d'un état à l'autre (Figure 1-2C).

Deux approches distinctes mais complémentaires permettent d'obtenir de l'information structurale sur les différents états des protéines : l'approche expérimentale et l'approche théorique. De manière générale, la structure la plus stable d'une protéine peut être déterminée expérimentalement à partir d'informations issues de spectres RMN, de patrons de diffraction de rayons X, et de patron de diffraction d'électrons. Puisque ces

^e Encombrement moléculaire.¹⁸

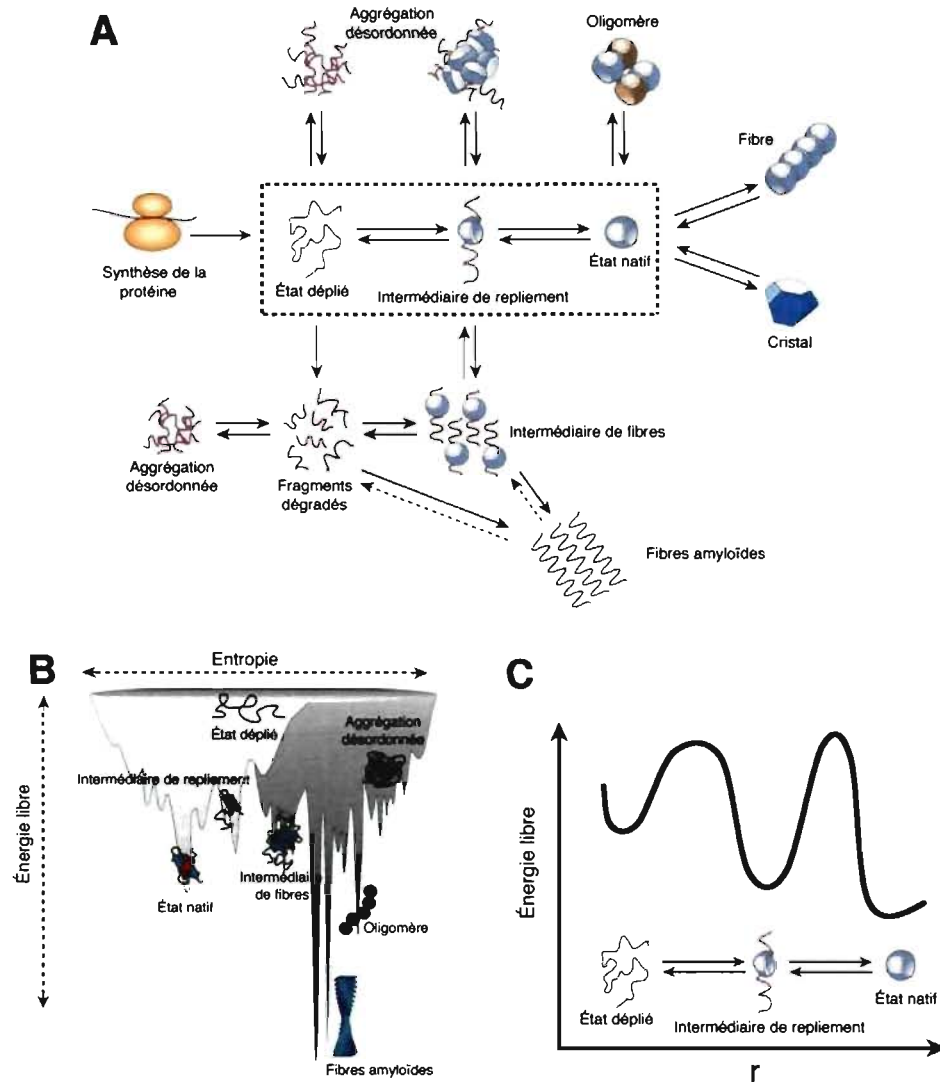


Figure 1-2. Les différents états d'une protéine et sa surface énergétique. A) Schéma représentant les états typiques empruntés par une protéine durant son existence (adapté d'un article de Vendruscolo & Dobson)²⁵. B) Exemple d'une surface énergétique de protéine (adapté de Jahn & Radford)²⁶. En principe, l'état déplié est composé d'une quantité astronomique de conformations (grande entropie); l'état natif de la protéine est généralement considéré comme l'état le plus stable (faible énergie libre). *Gris pâle*: états utiles de la protéine; *gris foncé*: états de la protéine nocifs pour l'organisme vivant. C) Détail de la surface énergétique du mécanisme de repliement d'une protéine (voir encadré dans A). Les minimums d'énergies correspondent aux états les plus stables (dépliés, intermédiaires et natif) et les maximums d'énergie correspondent aux états de transition.

méthodes requièrent des temps d'acquisitions relativement élevés, la caractérisation des structures transitoires de courte durée de vie ^f doit donc être réalisée à l'aide de stratégies indirectes ou à l'aide de méthodes spectroscopiques à plus faible résolution. L'approche expérimentale est très coûteuse en temps et en argent et est limitée à l'analyse de structures stables qui conservent leur intégrité dans les conditions expérimentales nécessaires à l'acquisition de données. Les approches théoriques actuellement en usage permettent pour leur part de modéliser l'état d'une protéine avec beaucoup de détails. Toutefois, ces simulations demeurent relativement limitées dans le temps étant donné la lourdeur des calculs nécessaires à la représentation du nombre astronomique de forces présentes au sein des protéines. D'une manière pratique, les simulations actuelles sur une protéine ne dépassent guère l'échelle des microsecondes en temps réel.²⁷ On est donc encore loin de pouvoir déterminer la structure des protéines à partir de leur séquence en acide aminé puisque le mécanisme de repliement des protéines se déroule généralement à une échelle de temps supérieur aux millisecondes. Toutefois, l'approche théorique permet d'explorer la surface énergétique d'une protéine en bordure d'une structure de protéine déterminée à partir de paramètres expérimentaux. Par exemple, le mécanisme de fonctionnement de certaines protéines a été simulé avec beaucoup de succès en performant des simulations de ces protéines à partir de l'état natif déterminé de manière expérimentale.^g En attendant que la puissance de calculs des ordinateurs soit grandement augmentée, l'avenir des simulations théoriques repose présentement sur la mise au point de modèles théoriques fiables et simplifiés.

^f Par exemple, les différents intermédiaires empruntés par une protéine au cours de son repliement ont généralement des durées de vie allant de quelques microsecondes à quelques secondes.

^g Les travaux du Professeur Martin Karplus de l'Université de Harvard à Cambridge sur la chaperonne d'*Escherichia coli* GroEL, et la protéine moteur F1-ATPase qui synthétise et hydrolyse de l'ATP représentent des exemples spectaculaires de réussites en la matière.²⁸

Ainsi, pour permettre une grande avancée dans le domaine d'étude des protéines, il faudra dans un avenir rapproché:

- 1) Mettre au point de nouvelles méthodes expérimentales qui permettent de caractériser efficacement les structures transitoires des protéines de courte demi-vie. Ce qui permettra de...
- 2) Générer davantage d'évidences expérimentales sur les mécanismes de repliement ou de fonctionnement des protéines qui serviront à tester et valider les modèles théoriques utilisés pour la simulation des protéines.

« Le travail rapporté dans cette thèse s'attaque donc au cœur du problème inhérent à l'étude des protéines en présentant une nouvelle méthode d'analyse expérimentale, la stratégie par sondage au tryptophane, rendant possible une percée majeure dans la caractérisation des structures transitoires de protéines. »

Le travail rapporté dans cette thèse s'attaque donc au cœur du problème inhérent à l'étude des protéines en présentant une nouvelle méthode d'analyse expérimentale, la stratégie par sondage au tryptophane, rendant possible une percée majeure dans la caractérisation des structures transitoires de protéines. Dans cette thèse, cette stratégie a été appliquée à l'étude des structures transitoires impliquées dans le mécanisme de repliement des protéines; elle aurait pu également servir à étudier tout autre état alternatif de protéines. Cette méthode permettra d'apporter un éclairage nouveau sur le mécanisme de repliement

des protéines tout en générant une nouvelle gamme de résultats expérimentaux qui contribueront grandement à tester et valider les simulations théoriques à venir.

Pour vous permettre de mettre mes résultats en perspective, les prochaines sections présenteront un bref aperçu de notre connaissance actuelle des mécanismes de repliement des protéines (1.1) en mettant une emphase particulière sur la limitation des méthodes expérimentales présentement en usage. Les bases théoriques de notre nouvelle méthode d'analyse seront ensuite exposées (1.2) suivie par la présentation de notre protéine modèle à l'étude, de nos hypothèses ainsi que de nos objectifs (1.3).

1.1. Le mécanisme de repliement des protéines

C'est en 1973 qu'Anfinsen démontre pour la première fois qu'une protéine dépliée^h détient toute l'information nécessaire pour lui permettre de retrouver rapidement sa forme nativeⁱ parmi un nombre astronomique d'autres conformations, sans la participation de facteurs extérieurs.³⁰ Du côté théorique, Levinthal démontre, quant à lui, dès 1968 que les protéines doivent bien se replier via des mécanismes précis car celles-ci prendraient une période presque égale à l'âge de l'univers pour retrouver leur structure la plus stable de manière aléatoire parmi la quantité astronomique de conformations leur étant accessibles.³¹ Dans les années 1970, les premières recherches sur le mécanisme de repliement des protéines ont tout d'abord porté sur les protéines qui pouvaient être facilement isolées et purifiées. Ces dernières, souvent relativement grosses (>150 acides aminés), affichaient généralement plusieurs transitions lorsque leur repliement était suivi à l'aide de méthodes spectroscopiques sensibles telles la fluorescence, ou le dichroïsme circulaire (Figure 1-3F).

^h La ribonucléase A et la nucléase de staphylocoque

ⁱ Certaines protéines nécessitent toutefois un coup de main additionnel pour y arriver via l'utilisation de prolyl isomérases ou de protéines chaperonnes.²⁹

Toutefois, les scientifiques ne tardèrent pas à découvrir que plusieurs de ces transitions représentaient en fait des populations de protéines se repliant plus lentement dû à la présence de facteurs ralentissant la progression de leur repliement (par exemple, la formation de ponts di-sulfures ou l'isomérisation de liens peptidiques *cis-trans* de l'acide aminé proline).³² De plus, on réalisa également que le repliement *in vitro* de grosses protéines pouvait fort probablement générer des états transitoires artefactuels puisque lors de ces expériences, toute la protéine se retrouve simultanément exposée à l'eau (Figure 1-2A, voir agrégation désordonnée).^j Ainsi donc, l'interprétation des différentes transitions détectées durant le repliement *in vitro* des protéines a rapidement dégénéré en controverse.

La mise au point des techniques de clonages d'ADN, d'expression recombinante et de mutagenèse dirigée a véritablement initié l'ère moderne de l'étude des mécanismes de repliement des protéines en rendant possible l'expression, la purification et le design de mutants pour n'importe quelle protéine dont la séquence d'acides aminés est connue (Figure 1-3A). Depuis lors, les scientifiques intéressés à la compréhension des lois fondamentales du mécanisme de repliement des protéines ont ainsi préféré porter leur attention sur des modèles plus simples, préférablement globulaires, sans pont di-sulfures ou acides aminés prolines. Ainsi, une des grandes avancées du domaine a été réalisée en 1991 lorsque Sophie Jackson et Alan Fersht ont mis à jour une protéine se repliant en l'absence apparente de structure intermédiaires transitoires c'est-à-dire via un mécanisme de repliement de type 2 états (seuls les états natifs et dénaturés étant détectés durant le repliement/dépliement de la protéine).³⁴

^j En comparaison, le repliement des protéines à l'intérieur de la cellule se produit plutôt au rythme de la synthèse de sa chaîne polypeptidique. Dans les cellules eucaryotes, la partie N-terminale est graduellement exposée au solvant, à la vitesse relativement lente de 4 acides aminés par minute.³³

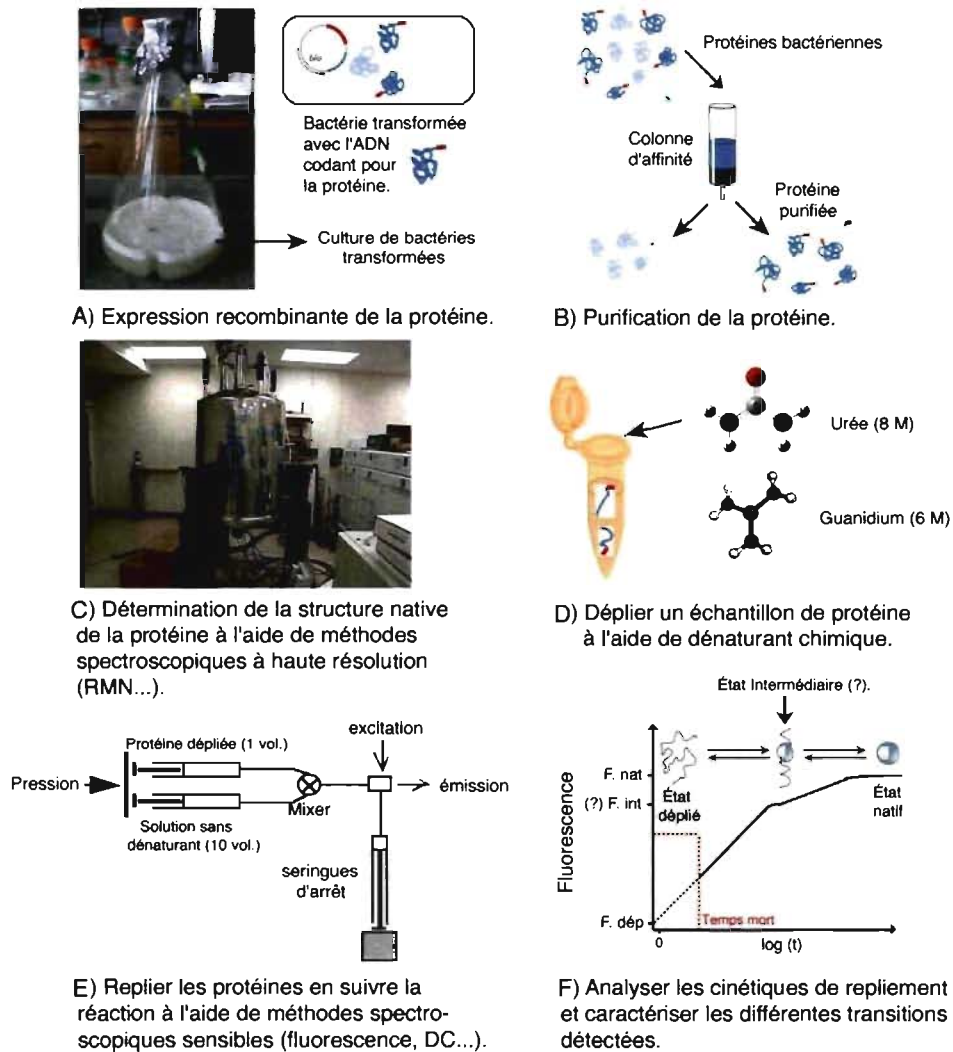


Figure 1-3. Comment étudier le mécanisme de repliement des protéines en six étapes. A) De nos jours, les protéines sont obtenues par expression recombinante dans des bactéries ou autres organismes. B) Les protéines et leurs mutants sont habituellement purifiés par colonne d'affinité en ajoutant une petite séquence d'acide aminé qui sera sélectivement accrochée et décrochée d'une colonne donnée. C) On détermine habituellement la structure native des protéines à l'aide de la spectroscopie RMN, de la cristallographie rayon-X ou de la microscopie électronique. D) Outre les dénaturants chimiques, on peut également dénaturer une protéine en utilisant des variations de pression ou de température. (E-F) L'analyse du repliement des protéines est généralement réalisée à l'aide d'un appareil de mixage combiné à un spectrophotomètre de fluorescence ou autres appareils de spectrophotométrie sensible.

Cette découverte a grandement contribué à simplifier l'étude du mécanisme de repliement des protéines car outre l'état natif et déplié, il ne restait qu'un seul état à caractériser pour déterminer le mécanisme de repliement de ces protéines: l'état de transition. Au cours des années 90, la majorité des études fondamentales sur le mécanisme de repliement des protéines ont ainsi porté sur la caractérisation de l'état de transition de petites protéines se repliant sans intermédiaire apparent.^{35,36} Toutefois plusieurs petites protéines ont également montré des signes de la présence d'intermédiaires et la caractérisation de ces derniers continue depuis lors à générer de nombreuses controverses.^{37,38} Au cours des prochaines sections, je présenterai un bref survol de notre connaissance actuelle des différents états empruntés par une protéine au cours de son repliement.

1.1.1. Les états natif et déplié d'une protéine

Une des principales caractéristiques des transitions de repliement et de dépliement des petites protéines naturelles est sans aucun doute leur étonnante coopérativité. En effet, les protéines naturelles ont été sélectionnées de manière à minimiser la stabilité de leurs états alternatifs: seuls les états déplié et natif étant habituellement assez stables pour être détectés à l'équilibre (Figure 1-4). On représente généralement l'état natif d'une protéine à l'aide d'une seule structure. Cette dernière permet habituellement d'optimiser l'enfouissement des acides aminés hydrophobes tout en maximisant le nombre d'interactions intramoléculaires : elle représente ainsi l'état le plus stable de la protéine (Figure 1-4A). Pour sa part, l'état déplié est, en principe, représenté par un ensemble hétérogène de conformations extrêmement disparates en équilibre rapide les unes avec les autres. Bien que ces deux états apparaissent comme étant très différents d'un point de vue structurel, ils possèdent néanmoins des énergies relativement similaires : la grande entropie

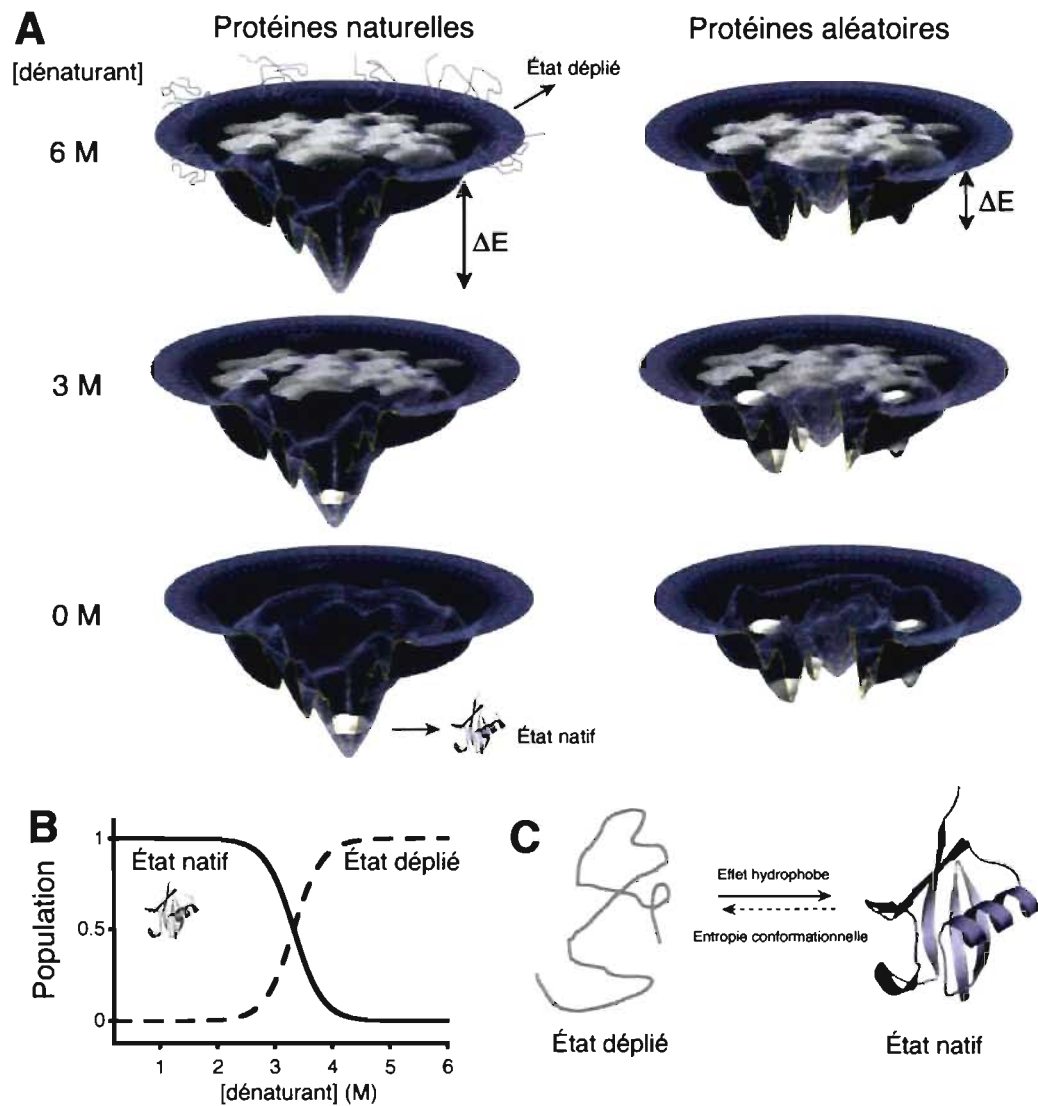


Figure 1-4. Coopérativité de la transition de repliement des petites protéines naturelles. A) Surface énergétique typique des petites protéines naturelles (*gauche*) et de peptides aléatoires (*droite*) (adapté d'un article de Wolynes).²⁷ L'espace conformationnel est représenté par la largeur tandis que l'énergie et la ressemblance à l'état natif est indiquée par la profondeur. Seuls les états natif (bas de l'entonnoir) et déplié (haut de l'entonnoir) sont généralement présents à l'équilibre pour les petites protéines naturelles. Les peptides aléatoires possèdent au contraire une surface énergétique très accidentée où plusieurs structures affichent des stabilités similaires (*droite*). B) De manière générale seuls les états natif et déplié sont détectés lors de l'ajout de dénaturant du à la forte coopérativité de la transition de repliement. C) Principales forces gouvernant l'équilibre entre les états natif et déplié.

conformationnelle de l'état déplié ^k faisant compétition avec l'énergie stabilisant l'état natif (en grande partie attribuable à l'enfouissement de ses résidus hydrophobes).^l C'est cette petite différence d'énergie qui existe entre l'énorme somme d'interactions qui stabilise chacun de ces deux états qui déterminera si une protéine se replie ou non.^m Cette raison explique sans doute pourquoi les potentiels utilisés lors des simulations de repliement de protéines ne sont pas encore assez précis pour permettre de déterminer avec succès l'état le plus stable qu'adoptera cette dernière.

Pour être efficace, une structure native de protéines doit donc contenir suffisamment d'interactions intramoléculaires pour permettre de déplacer l'équilibre du système de l'état déplié vers une seule conformation (Figure 1-4C). En plus, la séquence d'acides aminés devra également minimiser toute combinaison d'interactions qui permettrait de stabiliser des conformations non essentielles de la protéine (Figure 1-4A : voir surface énergétique des protéines aléatoires). Cette contrainte liée à l'optimisation du nombre d'interactions à l'intérieur même de la protéine fait en sorte que le nombre de structures qui permettent d'obtenir une stabilité suffisante s'avère relativement limité.ⁿ En revanche, un même type de plie peut facilement être obtenu à l'aide de plusieurs séquences d'acides aminés différentes car une certaine «redondance» chimique est observée parmi les vingt acides aminés. Par exemple, le repliement de type ubiquitine est une structure très utilisée par la nature : un peu plus de 1% des structures de protéines résolues jusqu'à présent affichent ce

^k Deuxième principe de thermodynamique : tout système tend vers le désordre maximal.

^l Les molécules d'eau s'ordonnent au contact de molécules hydrophobes ce qui contribue à diminuer l'entropie conformationnelle du système.

^m La fraction de protéine se repliant dans la forme native pouvant être déterminée par la distribution de Boltzmann: $K_e = \exp(-\Delta G_{U-N}/RT)$

ⁿ Pour quelques centaines de milliers de protéines, il y aurait seulement entre 1000 à 10 000 types de plies distincts.³⁹

type de plie malgré le fait que ces protéines aient des séquences d'acides aminés et des fonctions très variées.⁴⁰

1.1.1.1. Structure de l'état natif

Avant de s'intéresser aux mécanismes de repliement des protéines, il importe tout d'abord de présenter les forces et motifs qui permettent de stabiliser l'état natif d'une protéine. Une simple observation des structures natives des protéines naturelles permet d'en apprendre beaucoup à ce sujet (Figure 1-5). Ainsi, de manière générale, les structures natives des protéines:

- 1) Rassemblent les acides aminés hydrophobes à l'intérieur de la protéine (Figure 1-5A);
- 2) Optimisent les ponts hydrogènes intramoléculaires en utilisant deux types de structures dites secondaires : les hélices- α (Figure 1-5B) et les feuillets- β (Figure 1-5C);
- 3) Agencent les résidus chargés à l'intérieur de la protéine dans des ponts salins;
- 4) Tirent avantage de la propensité de certains acides aminés à se retrouver dans certaines conformations spécifiques (Figure 1-5D).

L'effet d'une mutation sur la stabilité d'une structure native dépendra donc autant de la position de l'acide aminé muté, que de la nature de l'acide aminé inséré à cette position. Par exemple, les acides aminés ayant des structures/fonctions similaires pourront habituellement être substitués l'un à l'autre sans affecter grandement la stabilité d'un état donné (voir Figure 1-5 : tables de propensité des acides aminés pour occuper différentes positions). En revanche, l'insertion d'acides aminés hydrophiles au cœur de la structure

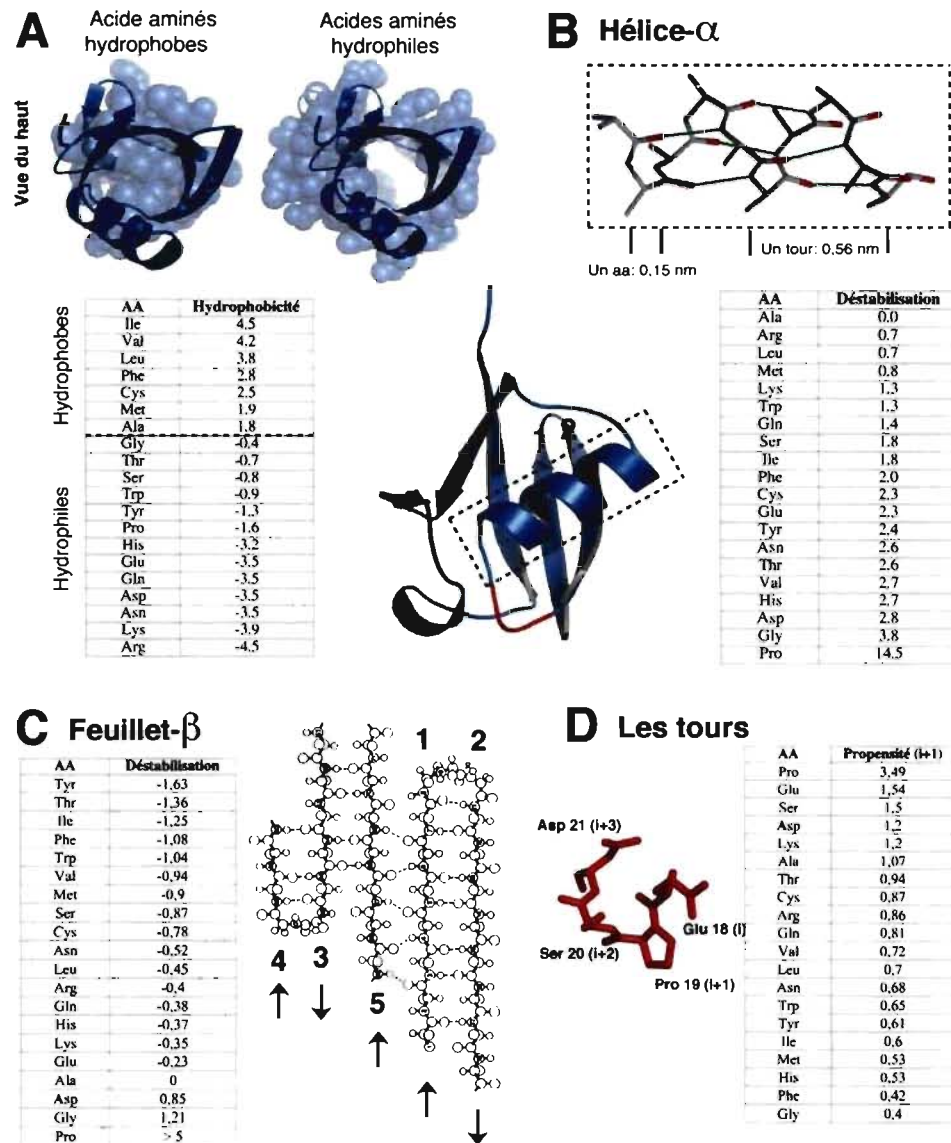


Figure 1-5. Forces et motifs permettant de stabiliser l'état natif des protéines et propension des acides aminés à se retrouver à certaines positions (ex : structure d'ubiquitine -1ubq).²² A) Les acides aminés hydrophobes d'une protéine sont généralement rassemblés au cœur de la structure. B-C) Les hélices- α (B), et les feuillets- β (C: adapté d'un article de Weber et coll.)⁹⁸ sont des structures secondaires qui permettent de maximiser le nombre de ponts H (B : lien vert; C : ligne pointillée) entre les protons amines et les groupes carbonyles de la chaîne principale. D) Différents types de tours permettent de relier les structures secondaires au sein d'une protéine. Le tour 18-21 d'Ubiquitine est un tour de type 1. La propension de chaque acide aminé à occuper l'une ou l'autre de ces positions est indiquée dans les différents tableaux.^{36,41,42}

hydrophobe de la protéine aura un effet beaucoup plus dramatique sur la stabilité de la forme native. Par exemple, deux mutations d'acides aminés hydrophobes en acides aminés hydrophiles dans le cœur d'une petite protéine globulaire suffiront habituellement pour rendre la structure native aussi stable que l'état dénaturé.^o Toutefois, beaucoup d'autres paramètres autres que la stabilité de la forme native entrent en jeu dans la sélection de chaque acide aminé des protéines naturelles. En effet, la séquence d'acides aminés doit également encoder le mécanisme de repliement, le mécanisme de fonctionnement et souvent même un mécanisme de régulation pour la fonction. De plus, cette séquence devra également permettre de minimiser la stabilité des nombreux états alternatifs possibles de la protéine qui pourraient nuire à son repliement (Figure 1-4A : minimiser les minimums locaux dans la surface énergétique de repliement) ou simplement interférer avec la fonction de milliers d'autres biomolécules présentes dans l'environnement direct de la protéine (Figure 1-2B : zone ombragée).

1.1.1.2. Structure de l'état déplié de la protéine

Notre connaissance des structures de l'état déplié d'une protéine demeure beaucoup plus sommaire que celle de l'état natif étant donné que la majorité des méthodes spectroscopiques en usage permettent uniquement de fournir des paramètres structuraux moyens pour une population de conformations. L'état déplié étant constitué d'une quantité potentiellement astronomique de conformations différentes, l'interprétation d'un paramètre expérimental moyen demeure ainsi grandement limitée. Bien que l'on représente généralement l'état déplié d'une protéine comme un ensemble de conformations de type

^o L'état natif d'une protéine globulaire d'environ 100 acides aminés est, de manière générale, environ 30 kJ/mol⁻¹ plus stable que l'état déplié. Le remplacement d'un acide aminé hydrophobe dans le cœur d'une protéine en acide aminé hydrophile entraîne généralement une déstabilisation supérieure à 15 kJ/mol⁻¹.³⁶

aléatoire (« random-coil »), plusieurs études expérimentales démontrent de plus en plus la présence de structures résiduelles au sein de cet état.⁴³ La caractérisation de ces structures, si sommaires soit-elles, est très importante car toute préférence conformationnelle de l'état déplié influencera grandement les étapes initiales menant au repliement correct ou incorrect de la protéine (Figure 1-2B). Au cours des dernières années, un nombre croissant de recherches ont également démontré que l'état déplié jouait un rôle beaucoup plus crucial au niveau du fonctionnement des mécanismes biochimiques que l'on ne l'avait d'abord imaginé. Par exemple, la transition de repliement du domaine liant les GTPases de la protéine Wasp est à la base d'un ingénieux mécanisme de régulation de l'activité de cette protéine.⁴⁴ Il semblerait même que plus du tiers des protéines encodées au niveau du génome d'un organisme pourrait bien demeurer désordonnées dans la cellule, c'est-à-dire ne pas se replier en structure tri-dimensionnelle fixe.⁴⁵

De manière générale, on considère que les forces ou interactions qui permettent de stabiliser la forme dépliée de la protéine sont de même nature que celles qui permettent de stabiliser l'état natif (Figure 1-5) à l'exception que les réseaux d'interactions intramoléculaires demeurent peu élaborés et très peu nombreux au niveau de cet état. Ainsi, parmi les différentes structures résiduelles détectées au niveau de l'état déplié, on retrouve des traces de structures secondaires, des interactions de type charge-charge, ainsi que des regroupements de résidus hydrophobes locaux et même distants (Figure 1-6B).^{43,46,47}

Approches permettant de déterminer la structure des états natif et déplié

La représentation des structures de protéines a subi une véritablement révolution depuis le premier modèle élaboré par Kendrew en 1957 à partir de plasticine (Figure 1-1A), jusqu'aux diverses représentations actuelles en trois dimensions en haute définition par ordinateur. Toutefois, au cours de toutes ces années, les différentes étapes menant à la

détermination de la structure d'une protéine dans un état donné sont restées relativement similaires:

- 1) Des résultats expérimentaux obtenus d'une technique donnée sont interprétés en termes de paramètres reliés à la structure moléculaire;
- 2) À partir des contraintes expérimentales obtenues à l'étape 1, un ou plusieurs modèles de protéine sont sélectionnés pour représenter la structure et l'énergie de cette molécule;
- 3) Une méthode d'optimisation est finalement utilisée de manière à minimiser les violations présentes au sein des conformations obtenues à l'étape 2.

Classiquement, les structures natives des protéines sont déterminées à partir de données expérimentales issues de patrons de diffraction de rayons X sur des cristaux de protéines, de spectres RMN, ou par microscopie électronique (Tableau 1). De manière générale, une seule structure est utilisée pour représenter l'état natif d'une protéine : celle qui permet d'intégrer le mieux toutes les contraintes expérimentales en minimisant le plus grand nombre de violations de structures covalentes. Toutefois, on commence de plus en plus à représenter l'état natif d'une protéine en intégrant plusieurs types de données expérimentales de manière à permettre une meilleure description de sa dynamique (Figure 1-6 *haut*). Par exemple, un nouveau type de simulation appelée DER (« Dynamic-ensemble refinement »),²⁵ effectuée sur l'état natif d'ubiquitine à partir de plusieurs contraintes expérimentales provenant de diverses techniques, a permis d'établir que la structure native de cette protéine devrait en fait être représentée par un ensemble de conformations distinctes en équilibre rapide ^p les unes avec les autres.⁴⁸ Ces simulations, à base de multiples contraintes expérimentales, ont de plus démontré que la structure native de cette

^p Le temps de vie des différentes conformations allant des picosecondes aux nanosecondes.

protéine peut également expérimenter des changements conformationnels beaucoup plus étendus grâce à de rares et larges fluctuations se produisant à une échelle de temps beaucoup plus longue. Cette description de l'état natif à l'aide d'un ensemble d'états permet ainsi d'obtenir une compréhension beaucoup plus précise de l'effet de la dynamique de la structure sur la fonction de la protéine, ainsi que sur la capacité de cette structure à réagir aux mutations.

La participation énergétique de chacun des acides aminés à la stabilité de l'état natif de la protéine peut également être déterminée de manière expérimentale à l'aide de la technique de mutagenèse dirigée et d'expériences de dénaturation. Par exemple, on peut déterminer la participation énergétique d'un acide aminé à une location spécifique dans la structure en mutant ce résidu en un autre acide aminé.^q À partir de courbes de dénaturation,^r et de test de fonctions, il sera alors possible de quantifier l'importance de la spécificité de cet acide aminé au niveau de la stabilisation de la structure native ainsi que son niveau d'implication dans le fonctionnement de la protéine.³⁶

^q Préférentiellement via des mutations de type conservatives i.e. qui permettent simplement de retirer un minimum d'atomes à l'acide aminé.

^r On dénature généralement une protéine avec l'aide de molécules dénaturantes tels le GdnHCl ou l'urée (Figure 3-4). Alternativement, on peut également dénaturer une protéine via un changement de température ou de pression.

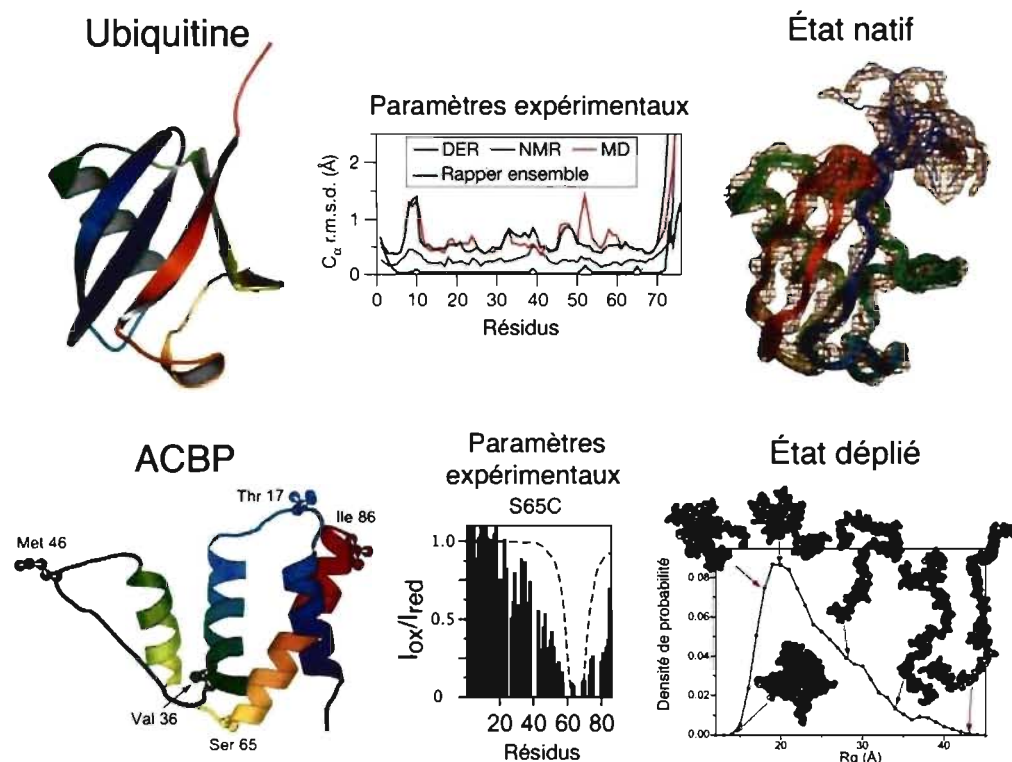


Figure 1-6. Ensembles de structures représentant les états natif et déplié de protéines obtenus en combinant des contraintes déterminées de manière expérimentales à des simulations de type « dynamic ensemble refinement » (« DER ») par ordinateur. A) L'état natif d'une protéine correspond à un ensemble de structures relativement similaires les unes aux autres. Ici, la chaîne principale ($C\alpha$) de la protéine ubiquitine est représenté à l'aide de 15 conformations obtenues à partir de simulations DER utilisant comme contraintes des paramètres issus d'expériences de RMN, de diffraction de rayons-X et de simulation de dynamique moléculaire (MD) (adapté d'un article de Lindorff-Larsen et coll.)⁴⁸. Cette représentation permet de rendre compte de la dynamique de la structure native. B) Des données de RMN de type « paramagnetic relaxation enhancement » obtenue sur l'état déplié de la protéine ACBP dans 1.6 M GdnHCl ont permis de générer plusieurs paramètres expérimentaux qui ont servis de contraintes pour simuler l'état déplié de cette protéine à partir de simulations DER (adapté d'un article de Lindorff-Larsen)⁴⁸. Chacune des 20 000 conformations générées est rapportée en fonction de son rayon de gyration (R_g). Les conformations les plus compactes comportent certaines interactions de type natif et non-natif (hélice- α et regroupement d'acides aminés hydrophobes distants) qui pourraient bien représenter les premières étapes du mécanisme de repliement de cette protéine.

Deux problèmes additionnels contribuent grandement à compliquer le travail des biochimistes lorsque vient le temps de déterminer les différentes conformations retrouvées au sein de l'état déplié d'une protéine. En effet, l'état déplié de la protéine est habituellement très peu présent dans des conditions physiologiques. Ainsi, pour parvenir à étudier ce dernier avec l'aide des techniques à haute résolution structurale telle la spectroscopie RMN, l'état natif de la protéine devra soit être déstabilisé à l'aide de mutations,⁴⁹ via l'ajout de dénaturant^{50,51} (voir Figure 1-4A: [dénaturant] = 6M), ou en augmentant la température du système.⁵² Un deuxième problème provient du fait que l'état déplié est beaucoup plus difficile à représenter que l'état natif étant donné son caractère hautement hétérogène. Toutefois, l'application des méthodes de simulation à partir de contraintes expérimentales ont récemment rendu possible une grande percée dans la représentation des différentes conformations présentes dans l'état déplié. Par exemple, à l'aide d'une méthode d'échantillonnage de structures de type « Monte-Carlo », l'équipe du Dr Vendruscolo est parvenue à générer plus de 20 000 conformations représentatives de l'état déplié de la protéine ACBP qui permettent de simuler de manière juste les paramètres issus d'expériences RMN (Figure 1-6 *bas*). L'analyse des conformations les plus compactes de cet ensemble montre la présence de certaines interactions de type natives et non-natives (hélice- α et regroupement d'acides aminés hydrophobes distants) qui pourraient bien représenter les premières étapes du mécanisme de repliement de cette protéine.⁴⁶

1.1.2. L'état de transition d'une réaction de repliement de type « deux états »

Comme nous l'avons mentionné dans la section précédente, la transition de repliement ou de dépliement d'une protéine implique une réorganisation complète de toutes les interactions stabilisant autant l'état natif que l'état déplié. Pour comprendre le

mécanisme par lequel cette transformation complexe est réalisée, il importe de caractériser les états intermédiaires empruntés par la protéine au cours de son repliement. Pour les mécanismes de repliement de type « deux états », il s'agit tout simplement de caractériser la structure de l'état de transition de cette réaction.

Les forces ou interactions qui permettent aux polypeptides naturels de retrouver rapidement leur structure native parmi une quantité astronomique d'autres structures potentielles sont en fait les mêmes qui permettent de stabiliser la structure native. Nous avons vu au cours de la section précédente que d'un point de vue thermodynamique,^s le design d'une structure efficace nécessite autant l'optimisation d'interactions de type natives^t que la minimisation d'interactions non-natives (Figure 1-4A : minimisation des **minimums locaux**). De manière analogue, il semble que d'un point de vue cinétique, la séquence d'acide aminé des protéines naturelles a été optimisée de manière à permettre aux interactions de type natives d'être formées plus rapidement que toutes autres interactions de type non-natives (Figure 1-4A; minimisation des **maximums locaux**). Les protéines naturelles arrivent donc à retrouver rapidement leur structure native puisque l'énorme bassin d'interactions de type natif accessible à l'état déplié va prendre le dessus sur tout autres combinaisons d'interactions non-natives. Ainsi, plus la conformation d'une molécule se rapprochera de la forme native, plus rapidement celle-ci sera stabilisée à nouveau par les interactions natives restant à former d'où cette forme d'entonnoir lisse pour représenter la surface énergétique des protéines naturelles.²⁷

Depuis plus d'une quinzaine d'années, l'état de transition de plus d'une trentaine de protéines se repliant via un mécanisme apparent à deux états a été mis à jour.^{36,53} De manière générale, il semble que la structure de cet état s'apparente grandement à la forme native de la protéine (même topologie) en beaucoup moins compact.⁵⁴⁻⁵⁶ Pour certaines

^s Qui est lié à la stabilité.

^t i.e. qui stabilisent la structure finale désirée

protéines (par exemple, le domaine SH3 de *src* Figure 1-7D), cet état apparaît plus polaire c'est-à-dire que les interactions impliquées dans la cinétique de repliement sont majoritairement concentrées dans une région spécifique de la protéine.⁵⁷ Un tel état de transition suggère que la protéine se replie suivant la formation d'un motif spécifique. Généralement, cette section de la protéine affichera une forte propension pour la conformation native (Figure 1-5). Par exemple, pour les petites protéines symétriques G et L, qui sont toutes deux formées de deux motifs de type épingle empaquetés sur une hélice- α ($\beta\beta\alpha\beta\beta$; topologie rappelant la protéine ubiquitine), la formation/empaquetage des différentes épingles sur l'hélice- α a lieu de manière séquentielle et semble régir par la différence de propensité entre les épingles: la première épingle à se former étant celle possédant la plus forte propensité.⁵⁸ Pour d'autres protéines, l'état de transition apparaît plutôt étendu (comme pour la protéine CI2),⁵⁹ c'est-à-dire qu'une grande proportion des interactions de type natives semble participer de manière partielle au mécanisme de repliement. Les protéines se repliant via un tel état de transition ne semblent donc pas nécessiter la formation d'un motif spécifique pour leur permettre d'amorcer leur repliement.

La structure de l'état de transition semble également déterminée par le type de repliement qu'adopte la protéine. Par exemple, les protéines avec des structures natives similaires semblent généralement se replier via des états de transition similaires (voir Figure 1-7C : valeurs Φ obtenues pour deux domaine SH3 avec faible identité de séquence).⁶⁰ Des simulations de repliement de protéines à partir de modèles d'acide aminés très simplifiés ont également réussi à prédire avec relativement de succès l'état de transition de plusieurs protéines.⁶¹⁻⁶³ Ces deux observations tendent ainsi à suggérer que le mécanisme de repliement des protéines est grandement dicté par le type de plie de leur structure native. Cette thèse est également appuyée par le fait que la vitesse de repliement

des petites protéines naturelles varie par plus de six ordres de grandeur en fonction de la complexité de la structure native ^u à atteindre.^{64,65}

Approches permettant de déterminer l'état de transition

L'approche permettant de déterminer les différentes conformations composant l'état de transition du mécanisme de repliement d'une protéine demeure sensiblement la même que celle utilisée dans le cas des états natif et déplié à l'exception que les paramètres expérimentaux utilisés comme contraintes proviennent uniquement d'expériences cinétiques (Tableau I). La participation énergétique d'un acide aminé à l'état de transition peut être évaluée de manière expérimentale en mutant cet acide aminé et en déterminant l'effet de cette mutation sur la cinétique de repliement ou de dépliement (Figure 1-7 A-C).³⁶ En effet, les vitesses de repliement et de dépliement d'une protéine se repliant via un mécanisme de type « deux états » sont données respectivement par les différences d'énergie libre entre l'état dénaturé et l'état de transition, et l'état natif et l'état de transition. Si la mutation d'un acide aminé affecte uniquement la vitesse de repliement de la protéine (Figure 1-7A *trace bleue*), on peut conclure que cet acide aminé est essentiel au mécanisme de repliement et que toutes ses interactions de type natives sont déjà formées au niveau de l'état de transition (valeur- $\Phi = 1$). Au contraire, si une mutation affecte simplement la vitesse de dépliement (Figure 1-7A *trace rouge*), alors l'acide aminé muté forme toutes ses interactions natives uniquement après l'état de transition et contribue ainsi seulement à la stabilisation de l'état natif (valeur- $\Phi = 0$).

^u La complexité d'une structure native peut être quantifiée par la distance moyenne entre les acides aminés entrants en interaction dans la structure.

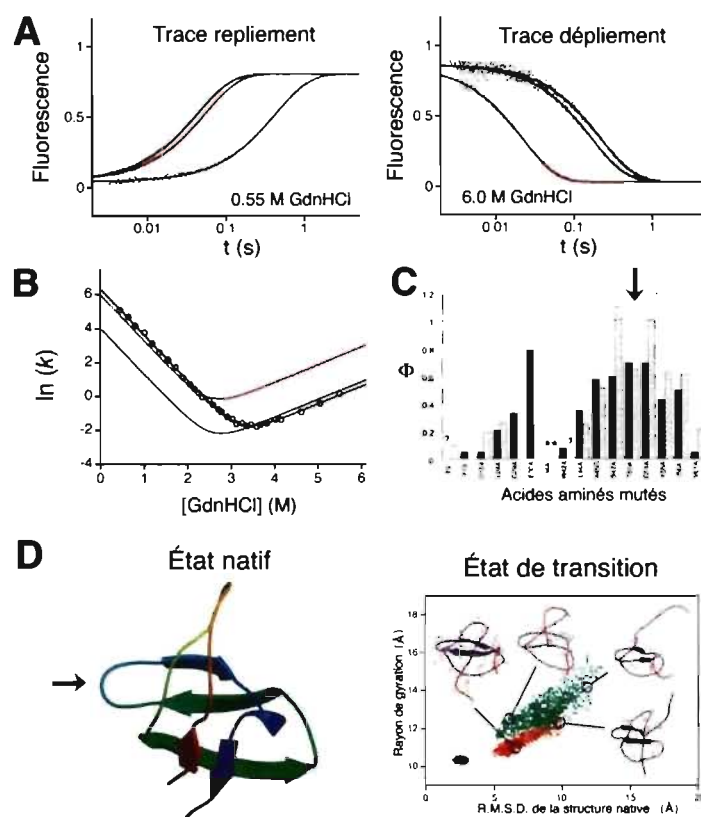


Figure 1-7. État de transition d'un mécanisme de repliement à « 2 états ». A) Traces de repliement (gauche) et traces de dépliement (droite) théoriques d'une protéine se repliant via un mécanisme de type « 2 états ». Les traces de repliement et de dépliement de mutants de la protéine dont l'acide aminé muté participe (bleu) ou ne participe pas (rouge) au mécanisme de repliement de la protéine sont également présentées. B) Vitesses de repliement et de dépliement théoriques obtenues à différentes concentrations de dénaturant pour les protéines *type sauvage* et mutantes décrites en A. Une valeur proche de 1, ou de 0, signifie respectivement que la mutation affecte uniquement la vitesse de repliement (participe au mécanisme de repliement), ou de dépliement (ne participe pas au mécanisme de repliement). C) Exemples de valeurs- Φ obtenues pour deux protéines de structures similaires (domaines SH3 : α -spectrin et Src) ayant peu d'identité de séquence (36%) (tiré d'un article de Martinez & Serrano)⁶⁰. D) Exemple d'un état de transition polaire (domaine SH3 de Src) déterminé à partir de contraintes fixées par les valeurs- Φ (tiré d'un article de Lindorff-Larsen et coll.)⁵⁶. L'épingle formée dans l'état de transition (bleu : valeurs- Φ proche de 1) est indiquée par une flèche dans les panneaux C et D. Les états de transitions (points rouges : 500 K; points verts : 640 K; différentes températures de simulations) sont rapportés en fonction de leur rayon de gyration et de leur degré de déviation de l'état natif (R.M.S.D.). Les points noirs représentent l'ensemble l'état natif.

Récemment, l'équipe du Dr Sosnick à l'Université de Chicago a également mis au point une méthode alternative permettant de sonder la présence d'interactions entre différentes positions au sein d'une protéine dans l'état de transition.⁶⁶ Cette méthode nommée valeur- Ψ requiert l'insertion de deux résidus histidines à des positions rapprochées dans la forme native (par exemple, aux positions i et $i+4$ sur une hélice- α ou sur deux brins- β interagissants). L'ajout de métaux divalents permet de stabiliser l'interaction entre les deux résidus histidine et de déterminer si les vitesses de repliement ou de dépliement sont ainsi affectées. Il est à noter que la quasi-totalité des études expérimentales portant sur la caractérisation des états de transition de protéines ont, à ce jour, été réalisées en déterminant la cinétique de repliement ou de dépliement de différents mutants en utilisant l'intensité de fluorescence d'un acide aminé tryptophane comme sonde (voir section 1.2).

1.1.3. Les intermédiaires de repliement

La mise à jour de nombreuses protéines, se repliant sans intermédiaire apparent,³⁶ a lancé un véritable débat concernant le rôle joué par les intermédiaires détectés lors du repliement des protéines.³⁸ D'un côté, il semble logique de croire que les intermédiaires sont essentiels au repliement des protéines car ils permettent de concevoir comment la protéine dépliée retrouve sa conformation native à partir d'une suite logique de transitions structurales simples. D'un autre côté, les multiples transitions détectées durant le repliement de certaines protéines^v peuvent également correspondre à des états mal repliés de la protéine retardant la progression du repliement (Figure 1-8A). Ainsi, le principal défi rencontré durant l'étude d'un mécanisme de repliement contenant plus de deux états consistera à déterminer la nature exacte des différentes transitions détectées. Par exemple :

^v D'autres signes permettent également de suspecter la présence d'intermédiaires (voir Figure 8).

1) Est-ce que les transitions détectées représentent la formation ou le réarrangement d'états intermédiaires participant au mécanisme de repliement (intermédiaires séquentiels), ou des transitions parallèles représentant la formation/réarrangement d'intermédiaires alternatifs (Figure 1-8A)?

2) Quelle est la structure des intermédiaires détectés? Est-ce qu'ils représentent des conformations mal repliées de la protéine, des conformations comportant une fraction des contacts natifs, ou des conformations comportant à la fois des contacts natifs et des contacts non-natifs?

Quelle que soit la nature des états intermédiaires détectés lors du repliement des protéines, leur caractérisation n'en demeure pas moins essentielle car elle permet de lever le voile sur tous les secrets entourant la mise au point d'une protéine au mécanisme de repliement efficace. En effet, il est aussi important pour un ingénieur de protéine de comprendre comment un état intermédiaire facilite le repliement d'une protéine, que de comprendre comment ce dernier peut également nuire à la progression du repliement. La caractérisation expérimentale juste et détaillée de tous types d'intermédiaires de repliement demeure également essentielle et primordiale pour tester et valider les simulations théoriques de protéines *ab initio*.

De manière générale, on peut regrouper les intermédiaires transitoires détectés lors du repliement des protéines en deux catégories. Les intermédiaires hâtifs (< 1 ms) sont habituellement détectés de manière indirecte sous la forme d'une amplitude manquante (« burst phase ») lorsqu'un instrument de mixage conventionnel de type « stopped-flow » est utilisé pour amorcer la réaction de repliement (transition ayant possiblement lieu durant le temps de mixage de l'appareil).^{68,69} L'étude de ces transitions nécessite généralement l'usage d'appareil de mixage haute vitesse.^{68,70} Pour leur part, la formation ainsi que le réarrangement des intermédiaires tardifs ont plutôt lieu à une échelle de

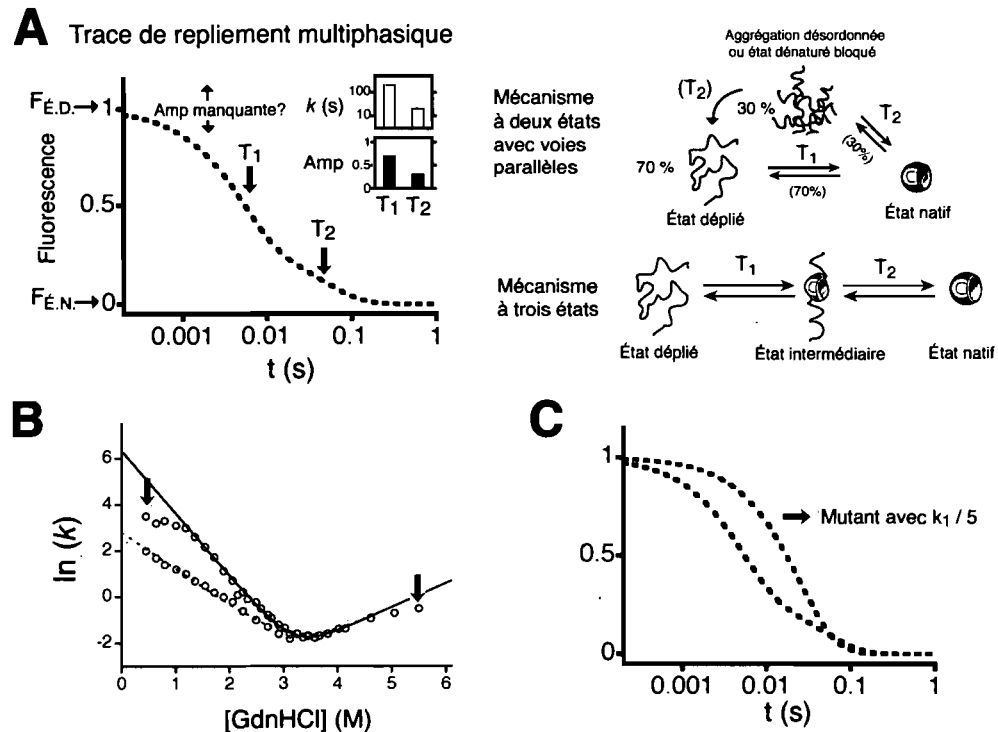


Figure 1-8. Les états transitoires détectés lors du repliement des protéines. A) Traces de repliement contenant deux transitions qui peuvent représenter des voies de repliement parallèles ou une population de protéines expérimentant deux transitions structurales successives. B) La présence d'intermédiaires de repliement peut également être détectée lorsque les vitesses de repliement ou de dépliement dévient d'un mécanisme apparent à deux états (courbe de chevron; voir Figure 1-7B). C) Limitation de la méthode de perturbation de structure pour étudier plus de deux transitions: la mutation d'un acide aminé qui réduit la vitesse de repliement de la première transition (T₁) entraîne facilement la disparition de la deuxième transition (T₂) lorsque les vitesses des deux transitions sont relativement rapprochées (<10x). D) L'utilisation de différentes méthodes spectroscopiques pour analyser le repliement permet d'obtenir de l'information concernant la nature d'une transition.⁶⁷ Toutefois, peu d'appareils de mixage sur le marché combinent l'utilisation de plus de deux méthodes spectroscopiques (habituellement fluorescence et DC).

temps moins rapide ($> 1\text{ms}$) et peuvent habituellement être détectés à l'aide d'un appareil de mixage conventionnel.

L'étude des intermédiaires hâtif est particulièrement pertinente car elle permet de déterminer si un type précis d'interactions a prépondérance sur les autres lors des premières étapes du repliement. En effet, plusieurs scientifiques soutiennent que le repliement d'une protéine procède d'abord via la condensation rapide de sa chaîne polypeptidique autour de ses acides aminés hydrophobes.^{71,72} Ce mécanisme, appelé « hydrophobic collapse », restreindrait ainsi la protéine dans un état transitoire compact qui faciliterait grandement la recherche de l'état natif. Le patron d'hydrophobicité de la chaîne polypeptidique serait donc en grande partie responsable de la détermination d'un tel mécanisme de repliement. Par opposition, certains croient plutôt que c'est la formation de structures secondaires, guidée par la propensité des acides aminés, qui dirigerait dans un premier temps la protéine vers un espace conformationnel plus rapproché de la structure native. Ce modèle de repliement nommé « framework », ou « diffusion/collision » propose plutôt que la structure native de la protéine proviendrait d'une collision effective entre ses différentes structures secondaires préalablement formées.^{73,74} Plusieurs évidences dans la littérature tendent à démontrer la présence de l'un ou l'autre de ces types d'interactions au premier instant du repliement mais le degré de participation de ces structures (surtout l'intermédiaire hâtif de type « hydrophobique ») au mécanisme de repliement demeure encore difficile à déterminer étant donné la trop courte demi-vie de ces structures. Par exemple, il se pourrait que l'état déplié d'une protéine adopte simplement une conformation plus compacte au moment d'être rapidement transféré dans des conditions natives sans que cette conformation lui permette toutefois de retrouver plus facilement la conformation native. Pour leur part, les intermédiaires tardifs de repliement représentent généralement des structures rapprochées de l'état natif qui nécessitent un réarrangement de motifs ou un empaquetage plus précis du cœur hydrophobe.

Approches permettant d'étudier les états intermédiaires

La principale difficulté rencontrée lors de l'analyse des états intermédiaires d'une réaction de repliement est sans aucun doute liée à leur très courte durée de vie. En effet, ces états sont généralement transitoires c'est-à-dire qu'ils ne sont pas assez stables pour pouvoir être détectés à l'équilibre. Les méthodes classiques permettant d'obtenir de l'information structurale à haute résolution ne peuvent donc pas être mises à profit car ces dernières demeurent trop peu sensibles pour permettre l'acquisition de données en un si court laps de temps.

Bien qu'il puisse paraître excitant, au premier abord, de détecter des intermédiaires transitoires durant le repliement d'une protéine, il est en revanche extrêmement difficile de démontrer si ces derniers participent ou non au mécanisme de repliement. En effet, une transition additionnelle peut également représenter une population distincte de protéines se repliant plus lentement due à la présence d'isomères *cis* ou *trans* non-natifs de liens peptidiques de résidus prolines⁷⁵ ou non-prolyl,⁷⁶ ou à une aggrégation désordonnée^{77,78} (Figure 1-8A : voies parallèles). Pour déterminer si une transition représente une voie parallèle, on peut toujours suivre le repliement de la protéine à partir de diverses sondes de manière à voir si cette transition affiche toujours la même amplitude.⁶⁷ Alternativement, des expériences de double saut « dépliement-repliement » rapide peuvent également permettre de vérifier si le repliement d'une population de protéines dépliées contenant uniquement des liens peptidiques dans l'isomère de type natif montrent également des transitions additionnelles. Enfin, il est également possible de muter tous les résidus prolines de la protéine et voir si les transitions additionnelles demeurent présentes. Si l'on croit qu'une transition ne semble pas être une voie parallèle, on doit alors prouver que cet intermédiaire participe bel et bien au mécanisme de repliement, c'est-à-dire que cet intermédiaire permet à la protéine de progresser dans son repliement. Pour en faire la démonstration, le

biochimiste devra déterminer que la vitesse de conversion de l'intermédiaire en l'état natif ou en une structure en aval de l'intermédiaire est plus rapide que la vitesse de dépliement de cet intermédiaire.³⁶

Tableau 1-1. Techniques expérimentales permettant d'obtenir de l'information structurale sur les états natifs et non-natifs des protéines.

États	Techniques
État natif	Spectroscopie RMN
	Cristallographie par rayon X
	Microscopie électronique
	Échange Hydrogène / Deutérium
	Mutagenèse dirigée
État déplié	Spectroscopie RMN
État de transition	Expériences cinétiques/mutagenèse dirigée :
	- Valeurs de Φ
	- Valeurs de Ψ
État intermédiaire	
i- > 1% à l'équilibre	Spectroscopie RMN
ii- transitoire	Valeurs de Φ (difficile)
	Échange rapide Hydrogène / Deutérium

Une fois la nature de l'intermédiaire caractérisée, on devra alors acquérir des données expérimentales permettant d'établir des contraintes structurales pour ce dernier. Puisque la demi-vie de l'intermédiaire est beaucoup trop courte pour permettre l'utilisation de méthodes à haute résolution (RMN), une des avenues à explorer consistera à tenter de

déstabiliser l'état natif de la protéine via des mutations spécifiques ^w de manière à rendre l'intermédiaire détectable dans des conditions natives.⁷⁹ Cette approche demeure en revanche très contraignante et aléatoire puisqu'elle nécessite beaucoup d'informations sur l'intermédiaire et/ou de tâtonnement pour déterminer la mutation à effectuer. De plus, l'intermédiaire détecté à l'équilibre devra également être caractérisé de nouveau pour montrer qu'il correspond bel et bien à une conformation participant au mécanisme de repliement.

Une stratégie, développée par l'équipe du professeur Roder, permet également de suivre la formation des ponts H durant le repliement de la protéine de manière indirecte à partir de la spectroscopie RMN.⁸⁰ Cette méthode consiste à effectuer plusieurs échanges rapides Hydrogène-Deutérium (« pulse H/D exchange ») à différents temps de repliement; les ponts H étant déjà formés ne permettant plus d'être échangés. Toutefois cette approche demeure difficile à réaliser et peu quantitative. Finalement, il est également possible d'utiliser la méthode des valeurs- Φ de manière à déterminer le degré de participation des interactions natives à la stabilisation de l'état intermédiaire. Cette approche, réalisée avec succès par l'équipe du Dr Radford sur la protéine IM7,⁸¹ nécessite toutefois que la vitesse de formation et de réarrangement de l'intermédiaire soit relativement différente (Figure 1-8D) et que le signal de fluorescence permette de distinguer clairement la population de l'intermédiaire dans les traces de repliement.

^w L'état natif peut également être déstabilisé de manière un peu moins sélective par des changements de température, de pression et de solvants.

1.2. Détection, caractérisation, et visualisation des structures transitoires de protéines par sondage au tryptophane

Comme nous l'avons vu au cours du précédent chapitre, il existe de bonnes approches expérimentales permettant d'obtenir de l'information structurale détaillée sur les états présents à l'équilibre tels l'état natif, l'état déplié, et les états intermédiaires relativement stables. De plus, les méthodes d'analyses par valeurs- Φ et Ψ représentent des stratégies complémentaires qui permettent de déterminer avec relativement de succès, le niveau d'implication des interactions natives dans l'état de transition de mécanismes de repliement de type « deux états ». Toutefois, nous avons aussi vu qu'il n'existait pas de méthode très efficace permettant d'obtenir de l'information sur la nature et la structure des états intermédiaires transitoires détectés durant le repliement de protéines étant donné leur très courte durée de vie.

Critères à considérer pour la mise au point d'une méthode de détection et de caractérisation efficace des intermédiaires transitoires

Plusieurs critères devraient entrer en ligne de compte pour concevoir une nouvelle approche qui permettrait de détecter et de caractériser les structures transitoires de protéines efficacement. Cette technique devrait entre autres:

- 1) Permettre une acquisition de donnée de manière très rapide c'est-à-dire beaucoup plus rapidement que la demi-vie typique d'un intermédiaire transitoire ($\ll \mu\text{s}$);
- 2) Être très sensible aux variations de structures;

- 3) Donner de l'information avec une résolution au niveau de chaque acide aminé;
- 4) Nécessiter une petite concentration de protéines de manière à minimiser le phénomène d'agrégation;
- 5) Être accessible à tous les laboratoires et peu dispendieuse.

1.2.1. La stratégie par sondage au tryptophane

Parmi toutes les sondes utilisées, la fluorescence de l'acide aminé tryptophane est sans aucun doute le signal spectroscopique le plus efficace pour détecter les transitions structurales de protéines (Figure 1-9).^{82, 83} Le principal atout de cette sonde réside dans le fait que l'intensité de fluorescence du tryptophane est extrêmement sensible aux détails chimiques de son environnement local (Figure 1-9D).^x Ainsi, un échantillon de protéine de moins de 10 μM , contenant un seul tryptophane, arrive à générer un signal de fluorescence avec un excellent ratio « signal/bruit » comparativement à la spectroscopie RMN qui requiert généralement l'usage d'échantillons de protéines dont la concentration est supérieure à 250 μM . Bien que les mécanismes à la base de la grande variation d'intensité de fluorescence du tryptophane dans les protéines naturelles ne soient pas encore entièrement élucidés (Figure 1-9D),⁸⁵ on constate généralement que l'intensité de son émission diminue lorsque le résidu est exposé à un environnement polaire (solvant) ou à des charges positives.^{86, y} Un autre avantage majeur du tryptophane est que sa fluorescence

^x L'intensité de fluorescence du tryptophane au sein des protéines naturelles varie de plus de 35 fois; les différences d'efficacité quantique allant de 0.35 à moins de 0.01.⁸⁴

^y Les charges négatives semblent, au contraire, induire une augmentation de la fluorescence du tryptophane lorsque présentes à proximité du fluorophore.

peut être détectée et intégrée très rapidement puisque la majorité des photons sont libérés moins de 10 ns suivant l'absorption. En plus d'être de loin le résidu le plus fluorescent des acides aminés (Figure 1-9C), le tryptophane absorbe également la majorité des photons émis par les deux autres acides aminés fluorescents (la tyrosine et la phénylalanine).⁸⁷ La fluorescence des protéines est donc grandement dominée par la fluorescence du tryptophane. Enfin, le tryptophane est le plus rare des acides aminés, représentant seulement 1/80^e des résidus présents dans les protéines (Figure 1-9E);⁸⁸ les petites protéines en possèdent donc généralement un seul, ou même aucun.

« Pourquoi alors ne pas simplement suivre la progression
du repliement d'une protéine sur toute sa surface en utilisant
le signal de fluorescence de résidus tryptophanes insérés à
plusieurs locations dans la protéine ? »

Pourquoi alors ne pas simplement suivre la progression du repliement d'une protéine sur toute sa surface en utilisant le signal de fluorescence de résidus tryptophanes insérés à plusieurs locations? En principe, l'insertion d'un résidu tryptophane dans la majorité des locations exposées au solvant d'une protéine (~1 acide aminé sur deux) ne devrait pas affecter sa structure de manière significative. De plus, la génération de plusieurs mutants tryptophanes d'une protéine est à la portée de tous les laboratoires tant par la simplicité des techniques nécessaires au design ^z que par son faible coût.

^z Clonage et mutagenèse dirigée.

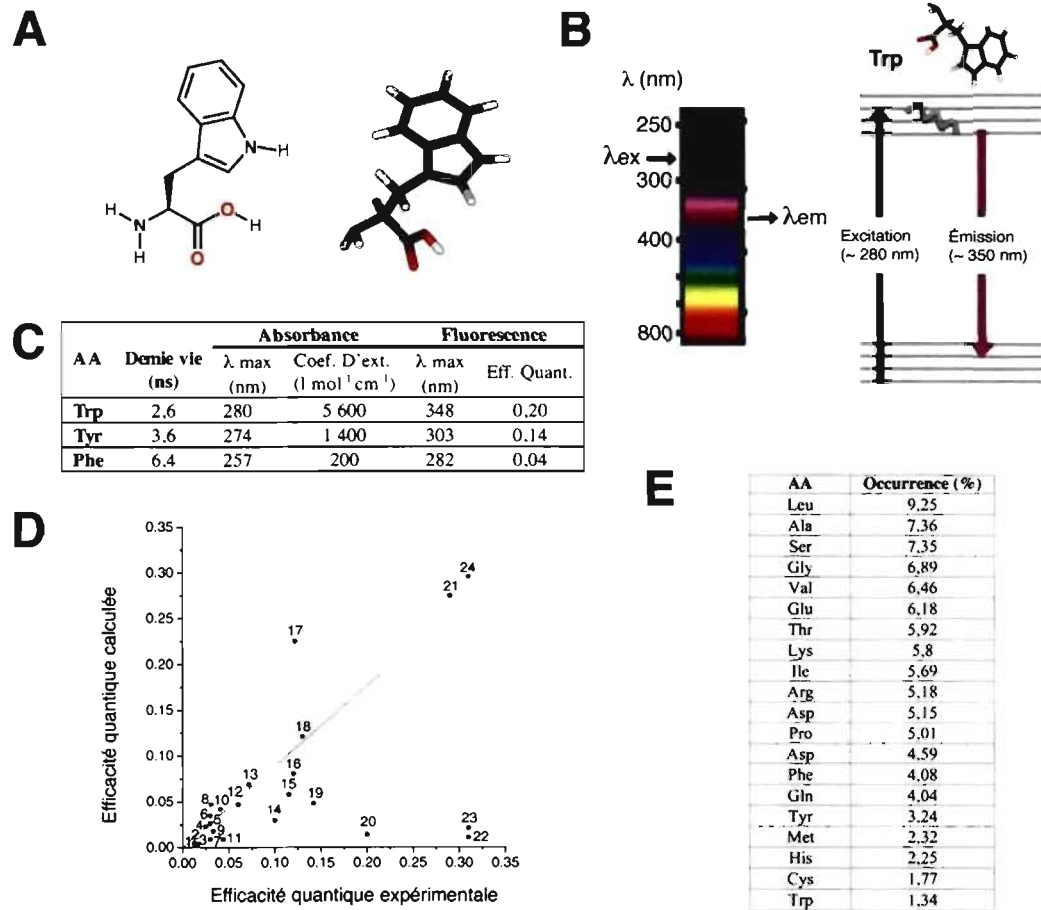


Figure 1-9. L'acide aminé tryptophane : structure et caractéristique de fluorescence. A) Structure du tryptophane en 2D (gauche) et en 3D (droite). B) Diagramme de Jablonski illustrant les transitions électroniques impliquées dans l'absorption et l'émission typique de photons du groupement indole du tryptophane. C) Paramètres caractéristiques des trois acides aminés fluorescents. Le tryptophane a la plus forte absorbance, le meilleur rendement quantique et absorbe généralement les longueurs d'ondes émises par la tyrosine et la phénylalanine. D) Efficacité quantique de fluorescence du tryptophane au sein de différentes protéines (tiré de Callis & Liu)⁸⁶. L'environnement chimique autour du tryptophane dans une protéine influence grandement son intensité de fluorescence (plus de 25 x). Bien que l'on arrive encore mal à prédire l'intensité de fluorescence d'un tryptophane en fonction de sa position dans une protéine (voir efficacité quantique calculée *versus* expérimentale), on admet généralement que plus celui-ci est exposé à un environnement polaire (solvant) ou à des charges négatives, plus sa fluorescence sera minimisée.⁸⁶ E) L'acide aminé tryptophane est le plus rare des acides aminés : en moyenne, 1 tryptophane par protéine de 75 acides aminés.⁸⁹

1.2.1.1. Utilisation de plusieurs mutants tryptophanes pour détecter et caractériser les états intermédiaires transitoires

En principe, la fluorescence d'un seul tryptophane ne devrait pas permettre la détection efficace de tous les changements conformationnels globaux et locaux expérimentés par une protéine au cours de son repliement. En effet, l'intensité de fluorescence du tryptophane ne devrait être sensible qu'aux changements structuraux prenant place dans son environnement local. Pourtant, la majorité des études expérimentales sur le mécanisme de repliement des protéines ne font usage que d'un seul tryptophane pour analyser le repliement ou dépliement de leur protéine.⁸³ Comment alors être assuré qu'une protéine se replie bien sans intermédiaire? De plus, comme nous l'avons vu plus haut, un des principaux problèmes liés à l'étude des intermédiaires transitoires de repliement consiste à déterminer si les transitions détectées durant le repliement représentent des événements séquentiels impliquant une seule population de protéines ou des transitions parallèles impliquant des populations distinctes de protéines (Figure 1-8A). En analysant le repliement d'une protéine à l'aide de tryptophanes insérés à différentes positions dans la protéine, il devrait ainsi être possible de détecter toutes les transitions structurales avec efficacité. De plus, les transitions représentant des voies de repliement parallèles pourront, en principe, être facilement identifiées puisque ces dernières devraient afficher des amplitudes relativement similaires indépendamment de la position de la sonde utilisée (Figure 1-10A).⁶⁷

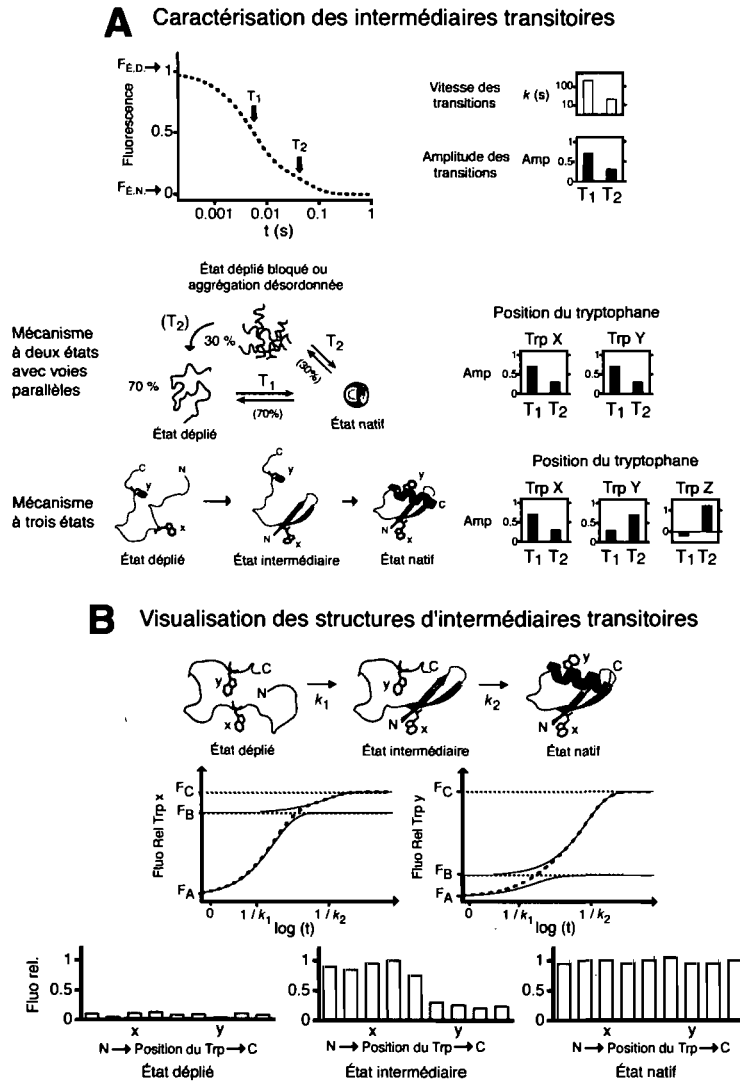


Figure 1-10. Caractérisation et visualisation des intermédiaires transitoires de repliement. A) Deux transitions représentant des voies parallèles vont montrer des amplitudes de repliement relativement similaires indépendamment de la position du tryptophane (Trp) utilisée pour suivre le repliement. Au contraire, des transitions représentant la formation d'un intermédiaire séquentiel vont montrer des amplitudes très différentes dépendamment de la position de la sonde. B) En principe, on pourrait caractériser la structure d'un intermédiaire transitoire en déterminant si la fluorescence d'un Trp dans cet état est de type natif ou déplié, à plusieurs positions. Par exemple, les Trp à la position x, et y devraient respectivement avoir une fluorescence de type natif (~ 1) et déplié (~ 0) dans l'intermédiaire.

1.2.1.2. Visualiser la structure d'un intermédiaire transitoire par sondage global au tryptophane

Notre compréhension actuelle des mécanismes régissant l'efficacité quantique de l'émission du tryptophane ne nous permet pas encore de pouvoir prédire avec exactitude l'intensité de fluorescence d'un tryptophane à partir des détails structuraux qui l'entourent (Figure 1-9D).⁸⁶ Il serait ainsi prématuré, à l'heure actuelle, de s'aventurer à prédire quelconques contraintes structurales aux abords d'un tryptophane en se basant uniquement sur ses caractéristiques de fluorescence. Toutefois, il serait en revanche possible de déterminer si l'environnement local en bordure d'un tryptophane dans un état intermédiaire est de type natif ou déplié en comparant simplement l'intensité de fluorescence du tryptophane obtenu dans chacun de ces trois états (Figure 1-10B). Par exemple, les tryptophanes localisés dans une région de la protéine qui serait repliée correctement (de manière native) au niveau de l'intermédiaire (Figure 1-10B : tryptophane à la position x), présenteraient des intensités de fluorescence de type natives dans l'intermédiaire. En revanche, des tryptophanes qui seraient localisés dans des régions encore dépliées de la protéine au niveau de l'intermédiaire (tryptophane à la position y) produiraient pour leur part une fluorescence de type déplié dans l'intermédiaire.

1.3. Protéines à l'étude, hypothèses et objectifs

Dans la présente étude, nous avons décidé d'appliquer notre nouvelle stratégie à l'étude du mécanisme de repliement de la protéine ubiquitine. La protéine ubiquitine ^{aa} porte bien son nom car on la retrouve aussi bien dans tous les types cellulaires eucaryotes que dans une grande proportion des laboratoires de recherche en biochimie. Sa grande

^{aa} Vient du latin *Ubique* qui signifie en toute chose et tout être.

popularité est en partie attribuable au fait que malgré sa petite taille, cette protéine se replie dans un état très structuré d'une remarquable stabilité^{bb} ce qui facilite grandement sa purification.^{91, cc} La protéine ubiquitine représente sans contredit le meilleur modèle pour la présente étude car :

1) Le repliement de cette protéine a été analysé à l'aide de toutes les méthodes d'analyse présentement en usage (voir sommaire des études Figure 1-11B) mais une controverse persiste toujours à savoir si un intermédiaire transitoire est détecté durant son repliement (Figure 1-11C);

2) Cette protéine possède les caractéristiques idéales pour des études expérimentales de repliement :

i- Elle est petite (76 acides aminés), globulaire, stable, et son état natif est très structuré;

ii- Elle est soluble à plus de 5 mM;

iii- Elle ne contient pas de ponts di-sulfures, pas de proline en conformation *cis* dans son état natif, pas de modifications post-traductionnelles qui viendraient compliquer l'analyse du repliement;

3) Cette protéine ne contient pas de tryptophane naturel.

^{bb} Cette particularité n'est probablement pas étrangère au fait qu'une des fonctions principales d'ubiquitine consiste à être attaché aux résidus lysines de protéines qui doivent être acheminées au protéasome pour fin de dégradation (voir Figure 1C).⁹⁰

^{cc} La protéine ubiquitine peut-être purifiée simplement en augmentant la température d'un échantillon de protéines aux environs de 90 à 95 °C là où la majorité des autres protéines cellulaires précipitent (ubiquitine restant repliée et soluble).

4) La structure de cette protéine est extrêmement utilisée par la nature : plus de 1% des structures de protéines résolues jusqu'à présent affichent ce plie ⁴⁰ malgré le fait que plusieurs d'entre elles montrent des séquences d'acides aminés et des fonctions très variées (ex : le DLR de *raf*, voir Figure 1-11A).^{dd}

Hypothèses et objectifs de la présente recherche

Le repliement de la protéine ubiquitine se produit via au moins quatre transitions tel que rapporté par la fluorescence d'un tryptophane inséré à la position 45 (Figure 1-11C).⁹⁴ Toutefois, seule la transition la plus rapide, T₁, a été caractérisée; cette dernière affichant les caractéristiques d'une transition de repliement de type « deux états ».^{93,95,96} Bien qu'il ait été proposé à de nombreuses reprises que les transitions plus lentes représentaient sans doute des petites populations de protéines se repliant plus lentement dû à la présence de résidus prolines fixés dans l'isomère *cis* ^{93,96,97} (voir Figure 1-8A : voies parallèles), la nature de ces transitions reste encore à déterminer.

Le premier objectif de ce travail a donc consisté à vérifier si des protéines qui possèdent la même structure qu'ubiquitine, mais des séquences d'acides aminés complètement différentes, se replient via un mécanisme similaire à ubiquitine comme ceci a été démontré pour d'autres structures de protéines ^{ee} (Chapitre 2).^{58,64} Pour cette portion de travail, le mécanisme de repliement de la protéine ubiquitine a donc été comparé au mécanisme de repliement du domaine liant ras (DLR) de la protéine *raf* (1rfa). Le DLR de *raf* est une protéine de 78 acides aminés affichant une structure remarquablement similaire

^{dd} Consultez à cet effet, l'excellente thèse de doctorat de mon distingué collègue le Dr. François-Xavier Campbell-Valois.

^{ee} Notamment pour des domaines SH3 appartenant à de nombreuses protéines (voir Figure 7C).

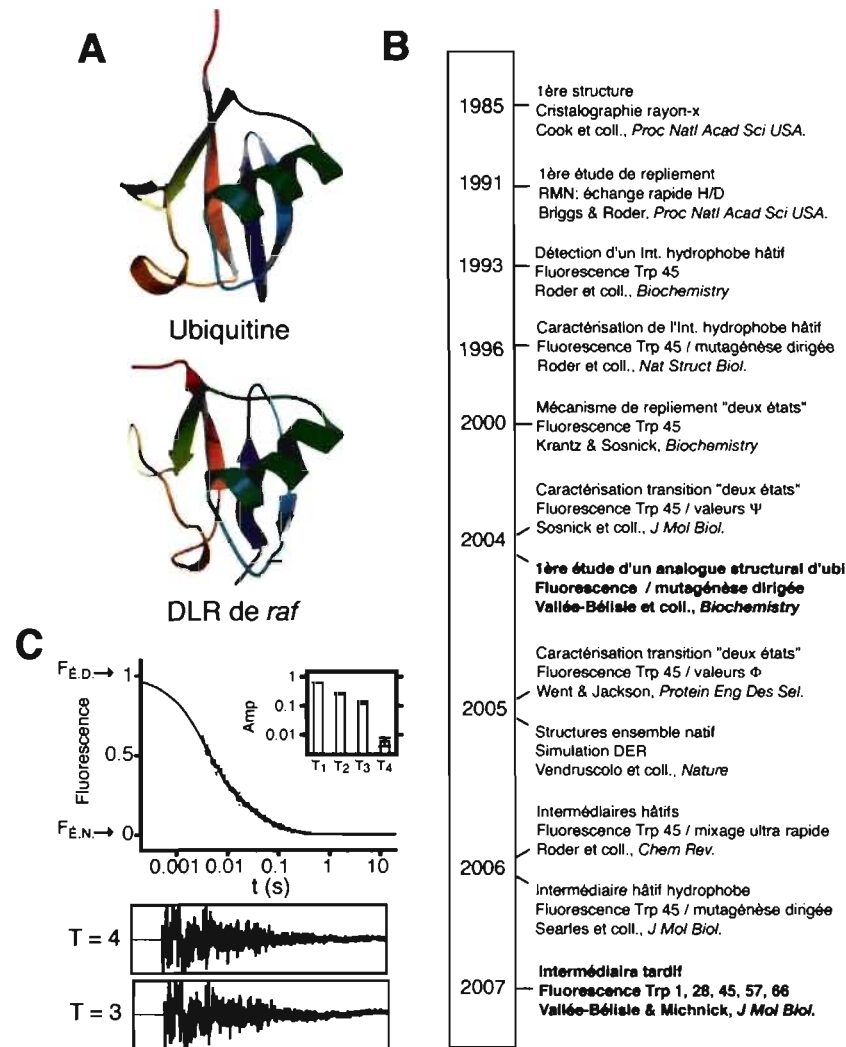


Figure 1-11. Ubiquitine : structure native, historique des principales découvertes et trace de repliement. A) Structure native d'ubiquitine (1ubq)²² et du domaine liant ras (DLR) de *raf* (1rfa)⁹², un analogue structural d'ubiquitine sans identité de séquence d'acides amines (11%). B) Résumé chronologique des principales études et modèles publiés sur le mécanisme de repliement de la protéine ubiquitine. C) Trace de repliement de la protéine ubiquitine obtenue en mesurant l'intensité de fluorescence d'un résidu tryptophane inséré à la position 45 dans la séquence. Le retour de la fluorescence de type native ($F_{E.N.}$) est réalisé via quatre transitions (voir résiduel des modélisations de la trace avec 3 transitions (T=3) et 4 transitions (t=4)). La transition T₁ affiche les caractéristiques d'une transition de type « deux états ».⁹³ La nature des transitions T₂, T₃ et T₄ reste à déterminer (sujet de cette thèse).

à celle d'ubiquitine sans toutefois démontrer la moindre identité au niveau de sa séquence en acide aminé (11% identité) (Figure 1-11A).⁹²

Le deuxième objectif de cette thèse a consisté à déterminer la nature des différentes transitions détectées durant le repliement d'ubiquitine (Figure 1-11C) en suivant le repliement de la protéine en utilisant la fluorescence de tryptophanes insérés à différentes locations sur sa surface –voir section 1.2.1.1. (Chapitre 3).

Finalement, le dernier objectif de cette thèse a consisté à mettre en application la méthode d'analyse par sondage global au tryptophane proposée à la section 1.2.1.2. pour visualiser la structure de l'intermédiaire transitoire détecté durant le repliement d'ubiquitine (Chapitre 4).

1.4. Références

1. Han, K. K. & Martinage, A. (1992). Post-translational chemical modification(s) of proteins. *Int J Biochem* **24**, 19-28.
2. Sidransky, D. & Hollstein, M. (1996). Clinical implications of the p53 gene. *Annu Rev Med* **47**, 285-301.
3. Harrison, R. S., Sharpe, P. C., Singh, Y. & Fairlie, D. P. (2007). Amyloid peptides and proteins in review. *Rev Physiol Biochem Pharmacol* **159**, 1-77.
4. Niemann, H. & Kues, W. A. (2007). Transgenic farm animals: an update. *Reprod Fertil Dev* **19**, 762-70.
5. Tyagi, A. K. & Khurana, J. P. (2003). Plant molecular biology and biotechnology research in the post-recombinant DNA era. *Adv Biochem Eng Biotechnol* **84**, 91-121.
6. Christou, P., Capell, T., Kohli, A., Gatehouse, J. A. & Gatehouse, A. M. (2006). Recent developments and future prospects in insect pest control in transgenic crops. *Trends Plant Sci* **11**, 302-8.
7. Barthelat, F. (2007). Biomimetics for next generation materials. *Philos Transact A Math Phys Eng Sci* **365**, 2907-19.

8. Vendrely, C. & Scheibel, T. (2007). Biotechnological production of spider-silk proteins enables new applications. *Macromol Biosci* **7**, 401-9.
9. Zhang, S. (2003). Fabrication of novel biomaterials through molecular self-assembly. *Nat Biotechnol* **21**, 1171-8.
10. Baez, J., Olsen, D. & Polarek, J. W. (2005). Recombinant microbial systems for the production of human collagen and gelatin. *Appl Microbiol Biotechnol* **69**, 245-52.
11. Merino, S. T. & Cherry, J. (2007). Progress and challenges in enzyme development for biomass utilization. *Adv Biochem Eng Biotechnol* **108**, 95-120.
12. Faber, K. & Kroutil, W. (2005). New enzymes for biotransformations. *Curr Opin Chem Biol* **9**, 181-7.
13. Ferrer, M., Golyshina, O., Beloqui, A. & Golyshin, P. N. (2007). Mining enzymes from extreme environments. *Curr Opin Microbiol* **10**, 207-14.
14. Lambrianou, A., Demin, S. & Hall, E. A. (2008). Protein engineering and electrochemical biosensors. *Adv Biochem Eng Biotechnol* **109**, 65-96.
15. Kohn, J. E. & Plaxco, K. W. (2005). Engineering a signal transduction mechanism for protein-based biosensors. *Proc Natl Acad Sci U S A* **102**, 10841-5.
16. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003). Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185-90.
17. Netzer, W. J. & Hartl, F. U. (1998). Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. *Trends Biochem Sci* **23**, 68-73.
18. Ellis, R. J. (2001). Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26**, 597-604.
19. Phillips, S. E. (1980). Structure and refinement of oxymyoglobin at 1.6 Å resolution. *J Mol Biol* **142**, 531-54.
20. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392-5.
21. Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-24.
22. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* **194**, 531-44.
23. Whitby, F. G., Masters, E. I., Kramer, L., Knowlton, J. R., Yao, Y., Wang, C. C. & Hill, C. P. (2000). Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408**, 115-20.

24. Yoshikawa, S., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., Yamashita, E., Inoue, N., Yao, M., Fei, M. J., Libeu, C. P., Mizushima, T., Yamaguchi, H., Tomizaki, T. & Tsukihara, T. (1998). Redox-coupled crystal structural changes in bovine heart cytochrome c oxidase. *Science* **280**, 1723-9.
25. Vendruscolo, M. & Dobson, C. M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. *Philos Transact A Math Phys Eng Sci* **363**, 433-50; discussion 450-2.
26. Jahn, T. R. & Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-17.
27. Wolynes, P. G. (2005). Energy landscapes and solved protein-folding problems. *Philos Transact A Math Phys Eng Sci* **363**, 453-64; discussion 464-7.
28. Karplus, M., Gao, Y. Q., Ma, J., van der Vaart, A. & Yang, W. (2005). Protein structural transitions and their functional role. *Philos Transact A Math Phys Eng Sci* **363**, 331-55; discussion 355-6.
29. Hartl, F. U. & Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-8.
30. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**, 223-30.
31. Levinthal, C. (1969). *Mossbauer Spectroscopy in Biological Systems*, Allerton House, Monticello, IL.
32. Baldwin, R. L. (1975). Intermediates in protein folding reactions and the mechanism of protein folding. *Annu Rev Biochem* **44**, 453-75.
33. Thulasiraman, V., Yang, C. F. & Frydman, J. (1999). In vivo newly translated polypeptides are sequestered in a protected folding environment. *Embo J* **18**, 85-95.
34. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30**, 10428-35.
35. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des* **3**, R81-91.
36. Fersht, A. R. (1999). *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*, Freeman, W.H., New-York.
37. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* **7**, 15-28.
38. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* **324**, 359-71.

39. Panchenko, A. R., Luthey-Schulten, Z., Cole, R. & Wolynes, P. G. (1997). The foldon universe: a survey of structural similarity and self-recognition of independently folding units. *J Mol Biol* **272**, 95-105.
40. Soding, J. & Lupas, A. N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays* **25**, 837-46.
41. Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-32.
42. Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci* **3**, 2207-16.
43. Bowler, B. E. (2007). Thermodynamics of protein denatured states. *Mol Biosyst* **3**, 88-99.
44. Kim, A. S., Kakalis, L. T., Abdul-Manan, N., Liu, G. A. & Rosen, M. K. (2000). Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature* **404**, 151-8.
45. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. & Obradovic, Z. (2001). Intrinsically disordered protein. *J Mol Graph Model* **19**, 26-59.
46. Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M. & Vendruscolo, M. (2004). Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J Am Chem Soc* **126**, 3291-9.
47. Cho, J. H., Sato, S., Horng, J. C., Anil, B. & Raleigh, D. P. (2008). Electrostatic interactions in the denatured state ensemble: their effect upon protein folding and protein stability. *Arch Biochem Biophys* **469**, 20-8.
48. Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M. & Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128-32.
49. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V. & Fersht, A. R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863-7.
50. Schwalbe, H., Fiebig, K. M., Buck, M., Jones, J. A., Grimshaw, S. B., Spencer, A., Glaser, S. J., Smith, L. J. & Dobson, C. M. (1997). Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry* **36**, 8977-91.

51. Freund, S. M., Wong, K. B. & Fersht, A. R. (1996). Initiation sites of protein folding by NMR analysis. *Proc Natl Acad Sci U S A* **93**, 10600-3.
52. Arcus, V. L., Vuilleumier, S., Freund, S. M., Bycroft, M. & Fersht, A. R. (1995). A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: implications for the initiation of protein folding. *J Mol Biol* **254**, 305-21.
53. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P., Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R. & Plaxco, K. W. (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* **14**, 602-16.
54. Fersht, A. R. (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc Natl Acad Sci U S A* **97**, 1525-9.
55. Pandit, A. D., Jha, A., Freed, K. F. & Sosnick, T. R. (2006). Small proteins fold through transition states with native-like topologies. *J Mol Biol* **361**, 755-70.
56. Lindorff-Larsen, K., Vendruscolo, M., Paci, E. & Dobson, C. M. (2004). Transition states for protein folding have native topologies despite high structural variability. *Nat Struct Mol Biol* **11**, 443-9.
57. Grantcharova, V. P., Riddle, D. S., Santiago, J. V. & Baker, D. (1998). Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol* **5**, 714-20.
58. Grantcharova, V., Alm, E. J., Baker, D. & Horwich, A. L. (2001). Mechanisms of protein folding. *Curr Opin Struct Biol* **11**, 70-82.
59. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* **254**, 260-88.
60. Martinez, J. C. & Serrano, L. (1999). The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* **6**, 1010-6.

61. Munoz, V. & Eaton, W. A. (1999). A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci U S A* **96**, 11311-6.
62. Galzitskaya, O. V. & Finkelstein, A. V. (1999). A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc Natl Acad Sci U S A* **96**, 11299-304.
63. Alm, E. & Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci U S A* **96**, 11305-10.
64. Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker, D. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* **39**, 11177-83.
65. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-94.
66. Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using psi analysis. *Chem Rev* **106**, 1862-76.
67. Wallace, L. A. & Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys Chem* **101-102**, 113-31.
68. Roder, H., Maki, K., Cheng, H. & Shastry, M. C. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15-27.
69. Myers, J. K. & Oas, T. G. (2002). Mechanism of fast protein folding. *Annu Rev Biochem* **71**, 783-815.
70. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
71. Ptitsyn, O. B. (1995). How the molten globule became. *Trends Biochem Sci* **20**, 376-9.
72. Kuwajima, K. (1989). The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* **6**, 87-103.
73. Bashford, D., Cohen, F. E., Karplus, M., Kuntz, I. D. & Weaver, D. L. (1988). Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* **4**, 211-27.
74. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci* **3**, 650-68.

75. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953-63.
76. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat Struct Biol* **8**, 452-8.
77. Silow, M., Tan, Y. J., Fersht, A. R. & Oliveberg, M. (1999). Formation of short-lived protein aggregates directly from the coil in two-state folding. *Biochemistry* **38**, 13006-12.
78. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc Natl Acad Sci U S A* **94**, 6084-6.
79. Sanz, J. M. & Fersht, A. R. (1993). Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry* **32**, 13584-92.
80. Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* **335**, 700-4.
81. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* **9**, 209-16.
82. Callis, P. R. (1997). 1La and 1Lb transitions of tryptophan: applications of theory and experimental observations to fluorescence of proteins. *Methods Enzymol* **278**, 113-50.
83. Royer, C. A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chem Rev* **106**, 1769-84.
84. Eftink, M. R. (1991). Fluorescence techniques for studying protein structure. *Methods Biochem Anal* **35**, 127-205.
85. Callis, P. R., Petrenko, A., Muino, P. L. & Tusell, J. R. (2007). Ab initio prediction of tryptophan fluorescence quenching by protein electric field enabled electron transfer. *J Phys Chem B* **111**, 10335-9.
86. Callis, P. R. & Liu, T. (2004). Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J Phys Chem B* **108**, 4248-4259.
87. Teale, F. W. & Weber, G. (1957). Ultraviolet fluorescence of the aromatic amino acids. *Biochem J* **65**, 476-82.
88. Klapper, M. H. (1977). The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem Biophys Res Commun* **78**, 1018-24.

89. Lehmann, W. D., Bohne, A. & von Der Lieth, C. W. (2000). The information encrypted in accurate peptide masses-improved protein identification and assistance in glycopeptide identification and characterization. *J Mass Spectrom* **35**, 1335-41.
90. Hochstrasser, M. (2006). Lingering mysteries of ubiquitin-chain assembly. *Cell* **124**, 27-34.
91. Jackson, S. E. (2006). Ubiquitin: a small protein folding paradigm. *Org Biomol Chem* **4**, 1845-53.
92. Emerson, S. D., Madison, V. S., Palermo, R. E., Waugh, D. S., Scheffler, J. E., Tsao, K. L., Kiefer, S. E., Liu, S. P. & Fry, D. C. (1995). Solution structure of the Ras-binding domain of c-Raf-1 and identification of its Ras interaction surface. *Biochemistry* **34**, 6911-8.
93. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696-701.
94. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054-63.
95. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* **337**, 463-75.
96. Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**, 229-37.
97. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* **3**, 193-205.
98. Weber, P. L., Brown, S. C. & Mueller, L. (1987). Sequential ¹H NMR assignments and secondary structure identification of human ubiquitin. *Biochemistry* **26**, 7282-90.

Chapitre 2. Article #1

Dans cet article, les mécanismes de repliement d'ubiquitine et du DLR de *raf* sont caractérisés à l'aide des méthodes d'analyse classiques (chapitre 1.1) de manière à valider l'hypothèse selon laquelle le mécanisme de repliement des protéines possédant des structures similaires serait le même. Faits intéressants : les traces de repliement des deux protéines comportent quatre transitions, la plus rapide de ces transitions représentant une transition de type «2 états». Ces travaux mettent toutefois en évidence les limites des approches traditionnelles qui jettent peu de lumière sur la nature des phases additionnelles observées dans le repliement des deux protéines (Figure 1-8).

Titre de l'article: *raf*RBD and Ubiquitin Proteins Share Similar Folds, Folding Rates and Mechanisms Despite Having Unrelated Amino Acid Sequences.

Auteurs: Alexis Vallée-Bélisle, Jean-François Turcotte, et Stephen W. Michnick

Publié dans: *Biochemistry* (2004) **43**, 8447-8458

Abbreviations

mUbi, mammalian ubiquitin; yUbi, yeast ubiquitin; *raf* RBD, Ras binding domain of the serine/threonine kinase *raf*; Gdn-HCl, guanidine hydrochloride; CD, circular dichroism.

2.1. Abstract

Recent experimental and theoretical studies in protein folding suggest that the rates and underlying mechanisms by which proteins attain the native state are largely determined by the topological complexity of a specific fold rather than by the fine details of the amino acid sequences. However, such arguments are based upon the examination of a limited number of protein folds. To test this view, we sought to investigate whether proteins belonging to the ubiquitin superfamily display similar folding behavior. To do so, we compared the folding-unfolding transitions of mammalian ubiquitin (mUbi) with those of its close yeast homologue (yUbi), and to those of the structurally related Ras binding domain (RBD) of the serine/threonine kinase *raf* that displays no apparent sequence homology with the ubiquitin family members. As demonstrated for mUbi [Krantz, B. A., and Sosnick, T. R. (2000) *Biochemistry* 39, 11696-11701], we show that a two-state transition model with no burst phase intermediate can describe folding of both yUbi and *raf* RBD. We further demonstrate that (1) all three proteins refold at rates that are within 1 order of magnitude (1800, 1100, and 370 s⁻¹ for mUbi, *raf* RBD, and yUbi, respectively), (2) both mUbi and *raf* RBD display similar refolding heterogeneity, and (3) the folding free energy barriers of both mUbi and *raf* RBD display a similar temperature dependence and sensitivity to a stabilizing agent or to mutations of a structurally equivalent central core residue. These findings are consistent with the view that rates and mechanisms for protein folding depend mostly on the complexity of the native structure topology rather than on the fine details of the amino acid sequence.

2.2. Introduction

During the past decade, simplifying concepts of protein folding behavior have emerged from the study of many single-domain, two-state folding proteins.¹⁻⁹ A key finding has been that the complexity of the native-state search (measured by the folding rate of a protein) is, surprisingly, correlated to a "contact order", a parameter which quantifies the average sequence separation that exists between residues making contact in the native structure of a protein.¹⁰ The existence of such a correlation between a parameter that describes the complexity of structural topologies and folding rate is supported by both experimental and theoretical evidence. Experimentally, folding rates of homologous proteins¹¹ and highly mutated variants^{12,13} were found to be relatively similar, and transition-state structures, characterized by Φ value analysis, of structurally homologous protein pairs have been found to be conserved.¹⁴⁻¹⁶ On a theoretical basis, several studies using low-resolution models with Gō-like approximations could be used to predict folding rates and the folding transition-state structure with good success.¹⁷⁻¹⁹ These results highlight the importance of native-state topology versus fine details of the polypeptide chain sequence as a determinant of protein folding rates and mechanisms.⁴

However, these arguments are built upon studies of a limited number of protein folds. Furthermore, folding studies comparing structurally related members often involved proteins, which have a discernible sequence homology (i.e., similarities in their fine sequence details).^{11-13, 15, 16, 21, 22} In the study presented here, we thus sought to examine whether folding rates and mechanisms of structurally related members belonging to the ubiquitin superfamily are similar in the absence of apparent amino acid sequence conservation. We chose the ubiquitin-like topology as a model for two reasons. First, the ubiquitin superfamily represents a remarkable example of structural versatility: its members display large sequence variation, various binding surface locations, and many unrelated functions.²³ In a recent version of the SCOP database (May 2003), where structures are

grouped by fold, superfamilies, and families, up to six sequence-distant families exhibit the characteristic ubiquitin-like secondary structure arrangement.²⁴ Second, the folding of mammalian ubiquitin (mUbi) has been studied extensively over the past decade, but the interpretation of its folding mechanism has remained the subject of controversy. In the mid-1990s, Roder and co-workers reported the existence of a putative "burst phase" intermediate in the refolding of mUbi^{F45W} at a low denaturant concentration (a transition thought to take place during the mixing period of the stopped-flow apparatus), which also appeared to be correlated with a rate constant rollover at 25 C.^{25, 26} The authors proposed that their results could be fit to a three-state model and interpreted as support for the existence of an early hydrophobic intermediate that is rapidly populated at the initiation of folding.²⁷ On the other hand, Krantz and Sosnick have recently demonstrated that a two-state model could adequately describe the folding mechanism for mUbi under conditions in which the fastest phase is not lost in the dead time of the stopped-flow apparatus.^{28, 29} Studies of additional ubiquitin superfamily members could thus aid in resolving this controversy.

For these reasons, we decided to compare the folding-unfolding pathway of mUbi to that of its close *Saccharomyces cerevisiae* yeast homologue, yUbi (sequence 96% identical), and, by contrast, to that of a structurally related protein displaying no apparent sequence homology with the ubiquitin proteins (Figure 2-1). Among the different superfamily members, *raf* RBD was identified as a good model for these studies because (1) its topology is virtually identical to that of ubiquitin, despite an apparent lack of sequence identity (<11%),^{30, 31} (2) it contains a single, partially buried tryptophan residue that could provide a fluorescent probe for monitoring the folding-unfolding transition (see Trp 114 in Figure 2-1), (3) it does not contain covalent disulfide bonds, and (4) it has two proline residues, Pro 64 and Pro 93, which are fixed in a *trans* conformation in the native structure of the protein, thus ensuring that only a small ratio of unfolded protein will be trapped in the slow rate-limiting prolyl isomerization at the initiation of refolding.³² In this work, we demonstrate that the folding-unfolding transition of the ubiquitin family and

superfamily members mUbi, yUbi, and *raf* RBD can be described well by an apparent two-state model, refold with similar rate constants, and show similar refolding heterogeneity. To further compare the folding mechanism of sequence-unrelated mUbi and *raf* RBD, we then present evidence to demonstrate that both proteins display equivalent folding free energy barrier variations when subjected to the same set of perturbations. Our results suggest that these proteins fold via similar folding pathways and thus reinforce the idea that the folding energy landscapes of proteins are relatively insensitive to the fine details of the amino acid sequences.

2.3. Materials and Methods

Protein Constructions and Expression. The yUbi coding sequence was amplified by a standard PCR procedure from the *SalI* engineered *Saccharomyces cerevisiae* UB14-derived Ubi gene³³ and cloned without further amino acid addition into a modified version of *Escherichia coli* expression plasmid pQE-32 (Qiagen). The mUbi coding sequence was obtained by (1) amplifying the yUbi cDNA using a 5' oligonucleotide which contains mUbi amino acid mutations (S19P, D22E, and S28A) and (2) cloning the PCR product into the yUbi expression vector using the *AgeI* site, which encodes amino acid positions 9 and 10 of the yUbi coding sequence. The Phe 45 Trp mutation was introduced into both ubiquitin coding sequences by replacing the wild-type (WT) nucleotides using a megaprimer PCR strategy.³⁴ The *raf* RBD coding sequence was amplified by a standard PCR procedure from the pWD333 plasmid,³⁵ and cloned into a modified version of pQE-32 with two extra amino acids added at the N-terminal extremity of the coding sequence (MG). Mutants L82I, L82V, and L82A were amplified from the parental sequence using a megaprimer PCR strategy³⁴ to replace the appropriate nucleotides and cloned back into the WT *raf* RBD expression vector. Coding sequence mUbi^{F45W} and *raf* RBD were alternatively cloned into a modified version of pQE-32 containing a polyhistidine tag. His₆-mUbi^{F45W} and His₆-*raf*

RBD thus contain the following amino acid insertions at their N-termini: MHHHHHHG and MHHHHHSMG, respectively. The coding sequences and promoter of the different expression vectors were verified by sequencing. Expression vectors were transformed into *E. coli* BL21 strain containing the *lacI* repressor in *trans* on a separate pREP4 plasmid. Cell cultures in ampicillin- and kanamycin-containing media were grown to an OD₆₀₀ of 0.8, and induced with 1 mM IPTG. After growing for 6 h, the cells were harvested by centrifugation and flash-frozen in liquid nitrogen.

Protein Purification. All proteins were resuspended in a minimal amount of buffer A [50 mM Tris-HCl (pH 8.0)] and sonicated on ice in 5 mL samples (~4% of buffer A per culture volume). His-tagged proteins (used for double-jump experiments only) were purified at a high concentration (~500 M) with Ni-NTA resin using the standard manufacturer's procedures (Qiagen). The eluted protein solution was then subjected to extensive dialysis to eliminate traces of imidazole. Purification of untagged proteins started with the addition of 50 mM MnCl₂ to precipitate DNA followed by a centrifugation to remove cell debris (these were recovered for mutant *raf* RBD L82A; see below). For the ubiquitin proteins, the supernatant was heated for 5 min at 80 °C and rapidly cooled, and the precipitated protein was removed by a second centrifugation. Purification was carried out with an 80% ammonium sulfate precipitation at 4 °C followed by size exclusion chromatography on a Sephadex G50 column (Pharmacia). The typical yield was 50 mg of pure protein/L as confirmed by Tricine SDS-PAGE. Purification of WT *raf* RBD and both mutants L82I and L82V was carried out by (1) applying the supernatant on a 15 mL cationic resin (S Sepharose from Pharmacia) equilibrated in buffer A (*raf* RBD pK_i ~ 8.9), (2) eluting *raf* RBD with 200-250 mM NaCl fractions using a NaCl gradient, (3) concentrating the protein by ammonium sulfate precipitate (80% at 4 °C) and resuspending it in a minimum amount of buffer B [50 mM Tris-HCl (pH 7.5) and 0.5 mM DTT], and (4) applying the concentrated *raf* RBD sample to a Sephadex G50 column (Pharmacia). Purification of mutant L82A was carried out by (1) washing cell debris with 50 mL of

buffer A and recovering the pellet from a second centrifugation, (2) resuspending this pellet in buffer C [6 M urea, 50 mM Tris-HCl (pH 8.0), and 0.5 mM DTT] and recovering the supernatant from a third centrifugation, (3) applying the supernatant on a 10 mL cationic resin (S Sepharose from Pharmacia) equilibrated in buffer C, (4) eluting *raf*RBD with 200-250 mM NaCl fractions using a NaCl gradient in buffer C, (5) reducing the urea concentration using dialysis against a 20× volume of buffer D [0.4 M Na₂SO₄ (enzyme grade from Fisher), 50 mM Tris-HCl (pH 7.5), and 0.5 mM DTT] and precipitating the L82A mutant with 80% ammonium sulfate at 4 °C, and (6) resuspending the protein pellet with buffer E [2.5 M Gdn-HCl (ultrapure grade from ICN), 50 mM Tris-HCl (pH 7.5), and 0.5 mM DTT] and applying it on a Sephadex G50 column (Pharmacia) equilibrated with buffer E. The fraction containing the protein was further subjected to dialysis using buffer B or buffer E for unfolding or folding experiments, respectively. The protein purity was confirmed with Tricine SDS-PAGE and the protein concentration later adjusted using Biomax-5 kDa Ultrafree centrifugal filter device (Millipore). The typical yield was 50 mg of pure protein/mL for WT *raf*RBD and mutants.

Data Collection. All experiments were performed in 25 mM sodium acetate buffer (pH 5.0) (ubiquitins) or in 50 mM Tris buffer (pH 7.5) with 0.5 mM DTT (*raf*RBD). The temperature was kept constant at 25 ± 0.1 °C unless otherwise specified. All measurements were taken using an Applied Photophysics SX18.MV stopped-flow fluorimeter with excitation wavelengths set to 285 nm (ubiquitins) or 281 nm (*raf*RBD) with a bandwidth of 2.5 nm. Fluorescence emission spectra were monitored by taking fluorescence measurements every 2 nm between 300 and 400 nm using a bandwidth of 5 nm, while equilibrium and kinetic denaturant dependence studies were performed by reading fluorescence intensity using a high-pass glass filter with 320 nm (ubiquitins) or 305 nm (*raf*RBD) cutoffs. All equilibrium stability curves were generated by using the equilibrium end point of 10 s for folding or unfolding kinetic traces. Three 10 s relaxation traces, each containing 1000 data points distributed in a logarithmic scale, were averaged for all kinetic experiments. Refolding experiments were carried out by mixing 1 volume of the denatured

protein (in 6.0 or 4.5 M Gdn-HCl solutions for ubiquitin proteins or *raf* RBD, respectively) with 10 volumes of solutions with denaturant concentrations ranging from 0 to 7.0 M. Unfolding experiments were performed by mixing 1 volume of native protein (in 1.0 or 0.5 M Gdn-HCl solutions for the ubiquitins or *raf* RBD, respectively) with 10 volumes of solutions with denaturant concentrations ranging from 7.0 to 0 M. Kinetic experiments performed on the *raf* RBD L82A mutant required different conditions because of its low stability. The unfolded sample only required a 2.0 M Gdn-HCl solution, while in the unfolding experiment, a 1:1 mixing was used to prevent aggregation by lowering the concentration of the protein sample. The final Gdn-HCl concentration of each experiment was determined by refractive index measurements with an Abbe 60 refractometer (Bellingham & Stanley Ltd.) using an independent set of dilutions. Final protein concentrations employed in single-jump experiments were set to approximately 20 or 15 M for the ubiquitins or *raf* RBD based on an extinction coefficient of 6990 or 5875 M⁻¹ at 280 nm, respectively (36). His-tagged proteins were used in double-jump experiments because they could be purified at the much higher concentrations required to achieve the 1:77 dilution of protein needed in these experiments. These began with rapidly unfolding 2 volumes of native His-tagged proteins (in 1.0 M and 0.5 M Gdn-HCl for mUbi and *raf* RBD, respectively) in 5 volumes of concentrated Gdn-HCl solutions (8.0 and 6.3 M Gdn-HCl for mUbi and *raf* RBD, respectively). After a 10 s delay, these samples were then subjected to rapid refolding using 1:10 mixing with low denaturant concentrations. Final protein concentrations employed in double-jump experiments were set to approximately 10 M.

Equilibrium analysis using a two-state model of denaturation, in which only the native and denatured states are populated at equilibrium (see eq 1), was applied to the entire set of fluorescence data extrapolated to infinity ($F_{\infty}^{\text{Gdn-HCl}}$) using the fit made on the kinetic traces in the denaturant unfolding and refolding studies (see Kinetic Analysis).

$$\text{Equation 1: } F_{\infty}^{\text{Gdn-HCl}} = [F_N^{\text{Gdn-HCl}} + F_U^{\text{Gdn-HCl}} \exp(m\text{Gdn-HCl} - \Delta G^{\text{H}_2\text{O}})/RT] / [1 + \exp(m\text{Gdn-HCl} - \Delta G^{\text{H}_2\text{O}})/RT]$$

Equation 1 assumes that the fluorescence intensity of the native and unfolded states ($F_N^{\text{Gdn-HCl}}$ and $F_U^{\text{Gdn-HCl}}$) are linear functions of Gdn-HCl concentration (Gdn-HCl). The model also assumes that the difference in free energy between the denatured and native states (G) varies linearly with the denaturant concentration, with a slope of m . $\Delta G^{\text{H}_2\text{O}}$ represents the difference in free energy between the two states in the absence of denaturant. The fraction of folded protein (F_N) was obtained from the raw stability curve data using the following transformation:

$$\text{Equation 2: } F_N = (F_{\infty}^{\text{Gdn-HCl}} - F_U^{\text{Gdn-HCl}} / \{1 + \exp[(m\text{Gdn-HCl} - \Delta G^{\text{H}_2\text{O}})/RT]\}) / F_N^{\text{Gdn-HCl}}$$

Kinetic Analysis. Following the dead time determination of the stopped-flow apparatus,³⁷ all the kinetic traces were transformed by adding 0.7 ms (proper setting of the initial reaction time) and removing all experimental data acquired before 3.5 ms. Kinetic parameters of the different traces were obtained by using the nonlinear regression analysis program Kaleidagraph (version 3.6, Synergy Software). All unfolding kinetic traces could be described well by either a single exponential or a constant function (unfolding experiments carried out under native conditions). Folding kinetic traces could require up to four exponential terms at low denaturant concentrations to be properly fit. Relaxation rates observed for the unfolding and the major refolding transitions were fit to a two-state model (chevron curve) according to equation 3.

$$\text{Equation 3: } \ln k_{\text{obs}} = \ln[k_F^{\text{H}_2\text{O}} \exp(-m_F \text{Gdn-HCl}) + k_U^{\text{H}_2\text{O}} \exp(-m_U \text{Gdn-HCl})]$$

The observed relaxation rate, k_{obs} , is the sum of the refolding and unfolding rates (k_F and k_U , respectively) at any Gdn-HCl concentration. The logarithms of k_F and k_U are assumed to be linear functions of the denaturant concentration with slopes m_F and m_U , respectively. $k_F^{\text{H}_2\text{O}}$ and $k_U^{\text{H}_2\text{O}}$ are the rates of folding and unfolding in the absence of denaturant, respectively. Standard deviations were obtained from the best fit of the data. To minimize the error in the fit, the rates of the fastest phase derived from the four-component fit at very low denaturant concentrations were not taken into account when comparing *raf* RBD chevron curves generated under different conditions (Figure 2-9A) or with different mutants (Figure 2-10B).

Intrinsic folding rates at 0.5 M Gdn-HCl (k_F 0.5 M), and intrinsic unfolding rates at 3.5 M Gdn-HCl (k_U 3.5 M), were obtained from a Chevron curve fit of the data points shifted to the left, such that k_F 0.5 M and k_U 3.5 M intersect the ordinate axis. This procedure was used in order to avoid propagated errors due to extrapolation of the data in absence of denaturant. $m\text{Ubi}^{\text{F45W}}$ values at 8 C were taken from the chevron curve reported in ref 25, while $m\text{Ubi}^{\text{F45W}}$ values measured in 0.4 M Na_2SO_4 and of the different mutants at position 26 were taken from ref 26. The two-state $k_F^{\text{H}_2\text{O}}$ values were not reported in the latter article because of the presence of a burst phase and rollover at low denaturant concentrations. However, on the basis of the recent demonstration that a two-state deviation in the folding of $m\text{Ubi}^{\text{F45W}}$ can be attributed to the disappearance of the fast two-state transition in the dead time of the stopped-flow experiment,²⁸ we extrapolated $k_F^{\text{H}_2\text{O}}$ from the kinetic m_F , m_U , and $k_U^{\text{H}_2\text{O}}$ values and from the equilibrium $C_{50\%}$ values (all determined in absence of the burst phase) using equation 4:

Equation 4:
$$k_F^{\text{H}_2\text{O}} = \exp[(RT \ln k_U^{\text{H}_2\text{O}} + m_U C_{50\%} + m_F C_{50\%})/RT]$$

Φ values were calculated from both the folding ($\Phi_F = \Delta\Delta G_{\ddagger-U} / \Delta\Delta G_{F-U}$) and unfolding experiments ($\Phi_U = 1 - \Phi_F = 1 - \Delta\Delta G_{\ddagger-F} / \Delta\Delta G_{F-U}$) by evaluating differences in the folding or unfolding free energy barrier obtained following mutation of WT amino acids to alanine. ΔG_{F-U} was determined from unfolding equilibrium experiments.

2.4. Results

Reversibility of the Folding-Unfolding Transition of yUbi and raf RBD. We first tested for the reversibility of folding and unfolding of both the Phe 45 Trp mutant of yUbi (yUbi^{F45W}) and raf RBD as previously demonstrated for the human ubiquitin (mUbi^{F45W}).²⁵ To do so, we compared the intrinsic fluorescence spectra and the unfolding free energies of the native and renatured states. Figure 2-2A shows that native (empty symbols) and renatured-state (filled symbols) fluorescence emission spectra are the same at a similar concentration with λ_{\max} values at around 340 and 330 nm for yUbi^{F45W} (squares) and raf RBD (circles), respectively. On the other hand, the denatured states (6.0 M Gdn-HCl) of both proteins (dotted symbols) display very similar fluorescence spectra (with λ_{\max} right shifted toward 350 nm) that are nearly identical to the fluorescence spectrum of free tryptophan in 6.0 M Gdn-HCl (×) at an equivalent concentration. The large change in fluorescence intensity observed following denaturation of both proteins indicated that intrinsic fluorescence would provide a sensitive probe for the folding-unfolding transitions for both yUbi^{F45W} and raf RBD. We generated equilibrium stability curves of native (filled symbols) and renatured (empty symbols) states of both proteins by monitoring changes in the fluorescence intensity versus Gdn-HCl concentration (Figure 2-2B). Similar, highly cooperative fluorescence intensity transitions were obtained with midpoint concentrations of denaturation ($C_{50\%}$) of 2.88 ± 0.01 and 2.93 ± 0.02 M for yUbi^{F45W} and 2.01 ± 0.03 and 2.03 ± 0.03 M for raf RBD for native and renatured states, respectively. A two-state

equilibrium modeling of the curves using nonlinear regression analysis of data fit to eq 1 (see Materials and Methods) revealed unfolding free energies in water ($\Delta G^{\text{H}_2\text{O}}$) of 6.4 ± 0.3 kcal/mol (native) and 6.8 ± 0.3 kcal/mol (renatured) for yUbi^{F45W} and 5.4 ± 0.7 kcal/mol (native) and 5.0 ± 0.6 kcal/mol (renatured) for rafRBD. These results show that native and renatured forms of both proteins display similar characteristics and further suggest that yUbi^{F45W} and rafRBD can be unfolded and refolded in a reversible manner.

Free Energy of the Folding-Unfolding Transition of the Ubiquitin Homologues and rafRBD. The equilibrium stability curves of yUbi^{F45W} (squares) and rafRBD (circles) were then transformed into the more convenient fraction folded representation shown in Figure 2-3A and compared to those for mUbi^{F45W}. As shown previously by CD equilibrium experiments,³⁸ the $C_{50\%}$ for yUbi^{F45W} was shifted to the left compared to that of mUbi^{F45W} (2.88 ± 0.01 and 3.78 ± 0.02 M Gdn-HCl, respectively; see Figure 2-3 and Table 2-1). However, despite this significant shift, the unfolding free energy of the yeast homologue extrapolated to 0 M denaturant ($\Delta G^{\text{H}_2\text{O}}$) was found to be only marginally smaller than that of mUbi^{F45W} (6.4 ± 0.3 and 7.0 ± 0.4 kcal/mol, respectively) as reflected by a larger determined m value for yUbi^{F45W} compared to that of mUbi^{F45W} (2.2 ± 0.1 and 1.80 ± 0.07 kcal mol⁻¹ M⁻¹, respectively; see Figure 2-3B). However, both ubiquitins were found to be more stable than the rafRBD (Figure 2-3B). The cooperativities of the transitions for both proteins, as quantified by m values, were similar but not identical within the experimental error for all three proteins (rafRBD having the highest value, 2.5 ± 0.3 kcal mol⁻¹ M⁻¹).

Folding and Unfolding Kinetics of the Ubiquitin Homologues and rafRBD. To characterize the kinetic behavior of mUbi^{F45W}, yUbi^{F45W}, and rafRBD, we analyzed the folding and unfolding relaxation following chemical denaturant jumps (Gdn-HCl) at 25 °C, again using the intrinsic tryptophan fluorescence as a probe. Unfolding traces could be fit to a single-exponential equation (data not shown), while up to four exponential terms were required to properly model the folding reactions. Figure 2-4 shows typical folding kinetic

traces obtained for mUbi^{F45W}, yUbi^{F45W}, and *raf* RBD at various Gdn-HCl concentrations (see Materials and Methods for details). Observed intrinsic folding and unfolding rate constants (empty and filled symbols, respectively) and the amplitudes of the different folding transitions are reported in Figure 2-6.

To assess whether a burst phase is observed within the mixing time of the stopped-flow apparatus following initiation of folding or unfolding experiments,²⁸ we then extrapolated the initial fluorescence values of the kinetic traces using the best-fit exponential or sum of exponential terms for each curve. Figure 2-5 presents both the extrapolated initial (circles) and final (squares) fluorescence intensities of all refolding and unfolding experiments performed on mUbi^{F45W}, yUbi^{F45W}, and *raf* RBD. For refolding experiments (empty symbols), initial fluorescence values (circles) for all three proteins showed no significant signs of deviation from the fluorescence intensity obtained for free tryptophan under similar conditions (×). For the unfolding experiments, the extrapolated initial fluorescence values (open circles) of all three proteins were not found to deviate from linearity. Final extrapolated fluorescence values obtained from both folding and unfolding experiments fit well to the equilibrium stability curve (solid line). Additional refolding experiments performed on *raf* RBD in the presence of 0.4 M Na₂SO₄ (a condition known to further stabilize hydrophobic interactions) did not reveal further deviation of the initial fluorescence intensity signal on traces measured at Gdn-HCl concentrations as low as 1.5 M [relaxation folding experiments performed at a lower Gdn-HCl concentration with 0.4 M Na₂SO₄ were too fast to be analyzed with our conventional stopped-flow apparatus]²⁸. These results demonstrate that for all three proteins, the intrinsic fluorescence amplitude going from denatured to native states, and vice versa, can be ascribed to a series of exponential components, suggesting that there is not a rapid, sub-millisecond burst-phase intermediate for the ubiquitins or for *raf* RBD.

Analysis of the refolding kinetics of both mUbi^{F45W} and yUbi^{F45W}, at low Gdn-HCl concentrations (<2.4 M for mUbi^{F45W} and <1.6 M for yUbi^{F45W}), shows that more than one exponential term is required to properly fit the traces (Figure 2-6). For mUbi^{F45W}, approximately 80% of the total refolding amplitude occurs in a fast two-state-like transition (open circles); two slower minor phases account for the remaining amplitude (squares and triangles; see Figure 2-6). As increasing amounts of Gdn-HCl reduce the rate of the predominant fast transition, both denaturant-independent minor phases are incorporated into the same unique transition. On the other hand, the folding of yUbi^{F45W} at low denaturant concentrations can be fit using the sum of only two exponentials (Figure 2-6): the middle transition observed in mUbi^{F45W} folding (squares) is probably already incorporated in the main folding transition of yUbi^{F45W}, the latter being 5 times slower in the yeast homologue (1800 vs 370 s⁻¹, respectively; see $k_F^{H_2O}$ in Table 2-1). For both ubiquitins, observed folding rates of the main transition (open circles) match the observed unfolding rates (closed circles) around the $C_{50\%}$ values determined from the equilibrium experiments. Subsequent fitting of these observed rates to a two-state chevron curve equation (see equation 3 in Materials and Methods) showed no apparent deviation from linearity (rollover) at low or high denaturant concentrations. Finally, the thermodynamic parameters obtained from the kinetic data (Table 2-1) correspond well to those obtained from equilibrium measurements (although the m value obtained from the equilibrium experiment on mUbi^{F45W} was found to be smaller than that obtained from kinetic data; see the Discussion), thus suggesting that the ubiquitin homologues fold and unfold via a unique transition-state ensemble (minor folding transitions not considered; see the Discussion).

raf RBD refolding at low denaturant concentrations (<0.6 M Gdn-HCl) displays a four-phase profile (Figure 2-6). At 0.41 M Gdn-HCl, the fastest phase accounts for around 60% of the refolding amplitude (open circles). It is followed by three additional transitions

accounting for ~30% (squares), ~3% (triangles), and 7% (×) of the total amplitude. As observed for mUbi, the folding rates of the additional transitions do not depend on denaturant concentration. The relaxation rate of the major folding transition (open circles) matches the unfolding relaxation rate (closed circles) around $C_{50\%}$ with no apparent deviation of folding and unfolding rates from the two-state chevron curve. Thermodynamic parameters obtained for *raf* RBD kinetic experiments (Table 2-1) are consistent with those obtained from the equilibrium measurements, thus suggesting that the folding and unfolding transition proceeds via a unique transition-state ensemble (minor folding transitions not considered; see the Discussion).

We then performed double-jump experiments to assess whether the additional transitions observed in mUbi and *raf* RBD refolding could be due to a slow proline or non-prolyl amino acid isomerization event.^{32, 39} Re-examination of mUbi refolding relaxation, acquired in oversampling mode (noise reduction strategy; compare mUbi refolding traces obtained in Figure 2-7 versus Figure 2-4), revealed a fourth slow minor phase with a very low amplitude (Figure 2-7, $T_4 \sim 2\%$). This four-phase refolding profile obtained for both proteins is well supported by the comparison of residuals obtained from fitting the traces to three or four exponential components (see the bottom of Figure 2-7). From the relative amplitude of the four-component fit, it is evident that the two fastest transitions observed for both mUbi and *raf* RBD refolding are not affected by the delay used to equilibrate the unfolded state (see the inset). However, the relative amplitude of the third transition for both proteins, and of the fourth transition for *raf* RBD, are significantly diminished when only 10 s is allowed for equilibration of the unfolded state after the denaturation of the native protein. Notably, all three additional transitions display similar rate constants for both proteins (see the inset, left).

Comparison of the chevron curves of all three proteins (Figure 2-8) reveals that the lower stability of *raf* RBD can be attributed to an increase of the unfolding rate [see extrapolated unfolding rates (- - -)] rather than a decrease in the folding rate. The folding rate of *raf* RBD (circles) was found to be even closer to that of mUbi^{F45W} (triangles) than that of yUbi^{F45W} (squares). On the other hand, the denaturant dependencies of the folding and unfolding rates for the ubiquitin homologues (m_F and m_U values, respectively) displayed similar values (see Table 2-1). One common interpretation of kinetic m values is that they measure the relative degree of compaction of the transition-state ensemble. From the following ratio, $m_F/(m_F + m_U)$ [called the β_T value]⁷, one can evaluate whether the transition state is more denatured-like or native-like. The β_T values of the ubiquitins were nearly identical (0.66 and 0.67 for mUbi^{F45W} and yUbi^{F45W}, respectively; see Table 2-1) but slightly more denatured-like than that of *raf* RBD (0.79). However, comparable variations in β_T are observed among members of protein families. For example, yeast, bovine, and rat ACBP have β_T values of 0.57, 0.61, and 0.69, respectively, despite the fact that they are homologues (sequence more than 50% identical)²²; cold shock proteins from *Bacillus caldolyticus*, *Bacillus subtilis*, and *Thermotoga maritima* have β_T values of 0.91, 0.93, and 0.86, respectively, despite their sequences being ~60% identical.¹¹

Folding and Unfolding Kinetic Variation of raf RBD following Various Perturbations. Having established that the general folding behavior of the ubiquitins and *raf* RBD is similar, we set out to probe the folding mechanism of *raf* RBD by performing perturbation studies that had previously been performed on mUbi^{F45W}. The folding-unfolding behavior of mUbi^{F45W} has been studied with respect to temperature changes, addition of a stabilizing agent, and mutations of a central core residue (Val 26).^{25, 26} We performed precisely the same studies on *raf* RBD. For the mutations, we chose the structurally equivalent position Leu 82 in *raf* RBD that corresponds to Val 26 in mUbi^{F45W}

(Figure 2-1). Figure 2-9 presents *raf* RBD equilibrium stability curves (panel A) and chevron curves (panel B) at 8 °C (circles) and with 0.4 M Na₂SO₄ (squares). Under both conditions, the observed folding and unfolding rate constants of *raf* RBD fit a chevron curve and the equilibrium parameters are well estimated by the kinetic data within experimental error (see Table 2-2). While addition of Na₂SO₄ increases both folding rates and stabilities of the folded forms (reduction of the unfolding rate), lowering the temperature led to a reduction of the folding rate, despite also increasing the unfolding free energy barrier. Figure 2-10 presents the equilibrium stability curves (panel A) and chevron curves (panel B) of *raf* RBD mutants L82A (circles), L82V (squares), and L82I (closed circles). All mutants display apparent two-state unfolding transitions with *m* values similar to that of the WT protein (see Table 2-2). From the equilibrium stability curves (Figure 2-10A), we observed that all mutants appear to be less stable than WT *raf* RBD (see Table 2-2). The reduction in stability for mutants L82I (closed circles) and L82V (squares) was caused by an increase in the unfolding rates, while the lower stability of L82A could be attributed to both a decrease and an increase of its folding and unfolding rates, respectively.

Figure 2-11 summarizes the effect of the different mutations performed at an equivalent central core residue on both *raf* RBD (gray) and mUbi^{F45W} (white) (see Materials and Methods for details on the extrapolation of the two-state folding transition rate of mUbi^{F45W} mutants). Free energy differences in the height of the folding (Figure 2-11A) and unfolding (Figure 2-11B) barriers obtained at 8 °C and with 0.4 M Na₂SO₄ are also shown. The key result revealed here is that folding free energy barriers for both *raf* RBD and mUbi^{F45W} are perturbed in the same manner (Figure 2-11A): (1) at lower temperatures ($\Delta\Delta G_{\ddagger-U}$ values obtained from the difference between $\Delta G_{\ddagger-U}$ at 8 °C and that at 25 °C are 1.45 ± 0.04 and 1.45 kcal/mol, respectively), (2) by the stabilizing agent ($\Delta\Delta G_{\ddagger-U}$ values obtained from the difference between $\Delta G_{\ddagger-U}$ in the presence and absence of

0.4 M Na₂SO₄ are -2.3 ± 0.1 and -2.1 kcal/mol, respectively), and (3) by mutations at a structurally equivalent position. For both proteins, the fastest folding mutant was the isoleucine substitution, followed by leucine, valine, and alanine, for the last of which folding was found to be even slower than in the WT proteins at 8 °C ($\Delta\Delta G_{\ddagger-U}$ for the mutation to alanine = 1.98 ± 0.07 and 2.0 kcal/mol for *raf* RBD and mUbi^{F45W}, respectively). Interestingly, faster folding mutants of both proteins are not the most stable; unfolding experiments reveal that the unfolding free energy barrier is maximal for the naturally selected amino acid (indicated by asterisks in Figure 2-11B). Northey and co-workers reported similar results in a study on an SH3 domain. They demonstrated that the folding rate could be significantly accelerated by replacing amino acids located at positions that are loosely packed in the transition state (Φ -value < 0.3) with those having bulkier side chains even though these mutations reduce the stability of the native state.⁴⁰

2.5. Discussion

Ubiquitin Homologues and raf RBD Fold and Unfold via an Apparent Two-State Mechanism. We have demonstrated that the folding and unfolding of ubiquitin homologues mUbi and yUbi and ubiquitin superfamily member *raf* RBD can be described well by an apparent two-state mechanism. This conclusion is supported by four observations: (1) the convergence of observed folding and unfolding rate constants around the equilibrium-determined $C_{50\%}$ (Figure 2-6, shoulder region), (2) the absence of rollover in the chevron curve at both low and high denaturant concentrations (Figures 2-6 and 2-7), (3) the absence of any relevant burst phase observed in the folding and unfolding process (Figure 2-5), and (4) the similar extrapolation of the thermodynamic parameters for the folding-unfolding transition from the kinetic and equilibrium experiments (Table 2-1). Thus, mUbi results obtained from our study are in good agreement with the recent study of Krantz and Sosnick²⁸ in which careful analysis of the fast refolding trace of mUbi^{F45W} acquired within

the detection limit of the stopped-flow apparatus shows no burst phase under conditions where an intermediate has been suggested to exist.^{25, 26} However, it is interesting to note that the equilibrium experiment performed with mUbi^{F45W} yielded smaller-than-expected m values (1.80 ± 0.07 and 1.9 ± 0.1 kcal mol⁻¹ M⁻¹ for the unfolding and refolding end point experiments, respectively, compared to 2.13 ± 0.04 kcal mol⁻¹ M⁻¹ for the kinetic experiment). This slight discrepancy was not observed for yUbi^{F45W} for which both folding (2.3 ± 0.1 kcal mol⁻¹ M⁻¹) and unfolding (2.2 ± 0.1 kcal mol⁻¹ M⁻¹) equilibrium-determined m values were similar to those extrapolated from kinetic data (2.15 ± 0.02 kcal mol⁻¹ M⁻¹). By using various probes, Khorasanizadeh and co-workers also obtained equilibrium-determined m values smaller than those determined from kinetic experiments (2.2 kcal mol⁻¹ M⁻¹)²⁵: 1.9 ± 0.1 kcal mol⁻¹ M⁻¹ using fluorescence at 353 nm, 1.9 ± 0.1 kcal mol⁻¹ M⁻¹ using NMR peaks of His 68 CH₂, and 2.0 ± 0.1 kcal mol⁻¹ M⁻¹ using molar ellipticity at 222 nm. On the other hand, Krantz and Sosnick found similar m values under the same conditions in both their equilibrium and kinetic experiments: 2.18 ± 0.02 kcal mol⁻¹ M⁻¹ (CD equilibrium experiments) and 2.17 ± 0.04 kcal mol⁻¹ M⁻¹, respectively.²⁸

Similar Folding Heterogeneity in the Refolding of the Ubiquitins, and raf RBD. As demonstrated earlier, the folding mechanism of raf RBD and yUbi^{F45W} can be described well as a two-state process when considering only the major fastest refolding transition. However, many additional minor phases are also observed during the refolding of both proteins at low denaturant concentrations. In their folding studies with mUbi^{F45W}, Khorasanizadeh and co-workers also observed two slower and smaller-amplitude transitions during the first 100 s of the folding relaxation traces.²⁶ For both transitions, rate constants decreased with an increase in the denaturant concentration (from 40 to 10 s⁻¹ and from 0.4 to 0.025 s⁻¹ for the middle and slowest transitions, respectively). In contrast to these results, Krantz and Sosnick reported from their two-state analysis that three to five

exponential terms were needed to properly fit the data over an extended time range.²⁸ In both reports, minor phases were proposed to be due to small populations of slowly refolding molecules with *cis*-prolyl peptide bonds preceding proline residues [mUbi contains three prolines in its coding sequence (Pro 19, Pro 37, and Pro 38)]. In the study presented here, four exponential terms were required to model the refolding kinetic traces of mUbi^{F45W} and *raf* RBD at low denaturant concentrations during the first 10 s (Figure 2-7). The additional transitions observed in both proteins were found to have similar rate constants and were relatively insensitive to Gdn-HCl [Figure 2-7 (inset, left) and Figure 2-6]. Furthermore, double-jump experiments suggest that the second transition observed during the refolding of both proteins does not display the characteristics of a prolyl *cis-trans* isomerization. Its amplitude is similar even when the protein is refolded from a nearly homogeneous population of all-*trans* amino acid conformations. However, the third transition in both proteins displays the characteristics of a slowly refolding population containing *cis* peptide bonds preceding non-prolyl residues. Its rate (~ 8 and ~ 7 s⁻¹ for mUbi and *raf* RBD, respectively), amplitude (4.5 and 7.1%, respectively), and dependence on equilibration time are similar to the behavior of a transition first described by Pappenberger and colleagues.³⁹ The fourth transition observed for *raf* RBD also appeared to be sensitive to equilibrium time, and displayed a rate and an amplitude that are characteristic of prolyl *cis-trans* isomerization.³² No such amplitude variation could be observed for the slowest transition of mUbi. However, it is relevant to note that this later phase remained hardly detectable because of both its low rate constant and relatively small amplitude. Finally, it is interesting to note that a previous pulse deuterium-hydrogen exchange experiment⁴¹ has already characterized three minor phases in the refolding of mUbi, which have rate constants similar to the three minor transitions identified in this study. In this study, amide hydrogen exchanges of residues 59, 61, and 69 were protected at a significantly slower rate (~ 45 s⁻¹) than the majority of the other amide protons, whereas amide protons of almost all amino acids also had $\sim 12\%$ of their complete amplitude protected at a rate of 7 s⁻¹. A fourth minor phase, accounting for $\sim 8\%$ of the complete protection amplitude for most protons,

was also identified with a rate constant as slow as 0.07 s^{-1} . More experiments are needed to determine whether the second transition observed in both mUbi and *raf* RBD refolding could represent a late intermediate accumulating along the folding pathway.²⁹

The Ubiquitins and raf RBD: The Same Fold and a Similar Folding Rate and Mechanism. As discussed by Plaxco and colleagues, no cases have been reported for which a homologous set of proteins exhibits folding rates that differ by more than 1 order of magnitude when their stabilities are taken into account.⁴ However, conservation of folding rate and mechanism within proteins displaying sequence homology could also be linked to the conservation of key folding residues.⁴² Thus, to further explore the extent to which folding behaviour of a protein is determined by the complexity of its fold rather than by the fine details of its sequence, kinetic studies of more structural homologues with unrelated sequences must be experimentally compared.

In the study presented here, we demonstrate that both structurally related *raf* RBD and the ubiquitin homologues fold and unfold via an apparent two-state model with relatively similar folding rates. In an attempt to compare the folding mechanism of sequence-unrelated *raf* RBD to that of mUbi, we then showed that very similar folding free energy variation was observed when both proteins are folded at lower temperatures, with added stabilizing agent, or when similar amino acids are inserted at a structurally equivalent position. Φ value analysis performed on the structurally equivalent position in both proteins yields similar medium values as determined from substituting WT amino acids with alanine (0.47 or 0.46 for *raf* RBD, 0.63 or 0.44 for mUbi^{F45W} using $\Delta\Delta G_{\ddagger-U}$ or $\Delta\Delta G_{\ddagger-F}$, respectively; see Table 2-2). Such values suggest that the central buried amino acid in the α -helices of both proteins displays approximately half of its native stability in the transition state. These observations suggest that, in addition to their similar folding rate constants, *raf* RBD and the ubiquitins also fold via similar mechanisms.

How Insensitive Are Folding Rates and Pathways to Fine Sequence Details? The results presented above provide a clear example of folding rate and mechanism conservation within ubiquitin superfold members. However, no apparent fine sequence detail conservation can be observed in the alignment of mUbi and *raf* RBD primary structure (Figure 2-1). Other comparative studies of structurally similar but not obviously homologous proteins have also led to similar conclusions. Chiti and co-workers showed that the transition states of the topologically related human muscle acylphosphatase (Acp) and activation domain of procarboxypeptidase A2 (ADA2h) have remarkably similar structures, despite their very low level of sequence identity (13%).¹⁴ The authors also showed that large folding rate differences between these proteins (0.24 and 900 s⁻¹, respectively) were consistent with their relative contact order values. Baker and co-workers elegantly showed that structurally related proteins L and G (sequences 15% identical) possess two similar and nearly isoenergetic folding pathways.⁴³⁻⁴⁵ Clarke and co-workers suggested that the folding pathway of Ig-like proteins shares common features, based on a strong correlation they observed between the folding rates and stabilities of many members with levels of sequence identity of less than 12%.⁴⁶ However, this study also illustrated the non-negligible contribution of fine sequence details (interactions defining the structure and stability of a fold) in determining the folding rates of the Ig-like proteins (FNfn9 and FNfn10 display 3 orders of magnitude of folding rate variation). Finally, Guerois and Serrano's research on Sso7d, a protein displaying the characteristic SH3 fold but lacking sequence identity with these other proteins, also suggests that not only the topology but also the particular sequence of a protein could sometimes be responsible for the selection of folding pathways (in the case of Sso7d, a glycine rich region located in the fifth α -strand).⁴⁷ Generalization of Φ value analysis to the entire sequence of both *raf* RBD and mUbi and the folding characterization of additional sequence-unrelated ubiquitin-like proteins will certainly provide further evidence of the relative insensitivity of folding rates and mechanisms of proteins to fine sequence.

Acknowledgment

The yUbi cDNA and *raf* RBD coding sequence were kindly provided by Dr. Nils Johnsson (Max-Delbruck-Laboratorium, Koln, Germany) and Dr. Julie E Scheffler (Bristol-Myers Squibb Pharmaceutical Research Institute, Princeton, NJ), respectively. We acknowledge François-Xavier Campbell-Valois and Dr. Melissa L. Lerch for critical reading of the manuscript, members of the Michnick group for helpful discussions, and professor Jeffrey W. Keillor for providing access to the stopped-flow apparatus.

2.6. Tables

Table 2-1. Folding Thermodynamic and Kinetic Parameters of mUbi^{F45W}, yUbi^{F45W}, and raf RBD at 25 °C Measured in Folding (f) and Unfolding (u) Experiments.

			$C_{50\%}$ ^a	m ^b	ΔG^{H_2O} ^c	RTm_F ^d	RTm_U ^d	$k_F^{H_2O}$ ^e	$k_U^{H_2O}$ ^e	β_T ^f
mUbi ^{F45W}	equil.	f	3.76 ± 0.02	1.9 ± 0.1	7.2 ± 0.4					
		u	3.78 ± 0.02	1.80 ± 0.07	7.0 ± 0.4					
	kinetic		3.79 ± 0.09	2.13 ± 0.04	8 ± 1	1.40 ± 0.03	0.73 ± 0.03	1800 ± 100	0.0021 ± 0.0006	0.66
yUbi ^{F45W}	equil.	f	2.93 ± 0.02	2.3 ± 0.1	6.8 ± 0.3					
		u	2.88 ± 0.01	2.2 ± 0.1	6.4 ± 0.3					
	kinetic		3.02 ± 0.04	2.15 ± 0.02	6.4 ± 0.4	1.45 ± 0.01	0.70 ± 0.01	370 ± 10	0.0078 ± 0.0008	0.67
raf RBD	equil.	f	2.03 ± 0.03	2.5 ± 0.3	5.0 ± 0.6					
		u	2.01 ± 0.03	2.7 ± 0.5	5.4 ± 0.7					
	kinetic		1.68 ± 0.08	2.75 ± 0.06	4.9 ± 0.4	2.17 ± 0.05	0.58 ± 0.01	1100 ± 100	0.31 ± 0.03	0.79

^a The $C_{50\%}$ values are in units of moles of Gdn-HCl per liter.

^b The m values are in units of kilocalories per mole per moles of Gdn-HCl per liter.

^c The free energies of folding (G) are in units of kilocalories per mole.

^d The kinetic m values for folding and unfolding (m_F and m_U , respectively) were multiplied by the factor RT (gas constant times the temperature) for the comparison with the equilibrium m values.

^e The rate constants are in units of inverse seconds.

^f The β_T were determined from the RTm_F/m_{kin} ratio.

Table 2-2. Folding Thermodynamic and Kinetic Parameters of *raf* RBD at 8 °C with 0.4 M Na₂SO₄ and of *raf* RBD with Different Central Core Residues at Position 82.

		<i>m</i> ^a	$\Delta G^{\text{H}_2\text{O}}$ ^b	RT <i>m</i> _F ^a	RT <i>m</i> _U ^a	<i>k</i> _F ^{0.5M} ^c	<i>k</i> _U ^{3.5M} ^c	$\Delta\Delta G_{\ddagger-U}^{\text{d}}$	$\Delta\Delta G_{\ddagger-F}^{\text{e}}$
25 °C	equil.	2.7 ± 0.5	5.4 ± 0.7						
	kinetic	2.51 ± 0.06	4.66 ± 0.08	1.92 ± 0.04	0.59 ± 0.01	148 ± 7	9.2 ± 0.2	-	-
8 °C	equil.	2.7 ± 0.3	5.3 ± 0.6						
	kinetic	2.50 ± 0.02	4.1 ± 0.1	2.01 ± 0.02	0.64 ± 0.01	14.9 ± 0.4	2.05 ± 0.03	1.45 ± 0.04	0.91 ± 0.02
Na ₂ SO ₄	equil.	3.3 ± 0.7	10 ± 2						
	kinetic	2.57 ± 0.08	7.5 ± 0.1	1.83 ± 0.05	0.73 ± 0.03	7000 ± 1000	3.9 ± 0.2	-2.3 ± 0.1	0.51 ± 0.04
L82I	equil.	2.3 ± 0.2	3.7 ± 0.4						
	kinetic	2.7 ± 0.1	3.6 ± 0.2	2.06 ± 0.09	0.63 ± 0.02	240 ± 20	90 ± 3	-0.29 ± 0.08	-1.35 ± 0.03
L82V	equil.	2.6 ± 0.4	3.6 ± 0.6						
	kinetic	2.52 ± 0.08	2.7 ± 0.2	1.83 ± 0.07	0.69 ± 0.01	88 ± 3	160 ± 5	0.31 ± 0.05	-1.69 ± 0.03
L82A	equil.	2.55 ± 0.08	1.15 ± 0.03						
	kinetic	2.6 ± 0.2	0.4 ± 0.2	1.7 ± 0.1	0.89 ± 0.02	5.2 ± 0.4	440 ± 30	1.98 ± 0.07	-2.29 ± 0.05

^a The *m* values are in units of kilocalories per mole per moles of Gdn-HCl per liter.

^b The free energies of folding (*G*) are in units of kilocalories per mole. Kinetic ΔG values were determined by adding $\Delta\Delta G_{\ddagger-F}$ and subtracting $\Delta\Delta G_{\ddagger-U}$ from the WT kinetic ΔG .

^c The intrinsic rate constants are in units of inverse seconds. To avoid error propagation resulting from extrapolation, the rate of folding (*k*_F) is reported at 0.5 M Gdn-HCl and the rate of unfolding (*k*_U) is reported at 3.5 M Gdn-HCl.

^d $\Delta\Delta G_{\ddagger-U}$ with a mutation or change in conditions (*x*) = $(RT \ln k_F^{0.5M})^{\text{wt}(25^\circ\text{C})} - (RT \ln k_F^{0.5M})^x$.

^e $\Delta\Delta G_{\ddagger-F}$ with a mutation or change in conditions (*x*) = $(RT \ln k_U^{3.5M})^{\text{wt}(25^\circ\text{C})} - (RT \ln k_U^{3.5M})^x$.

2.7. Figures

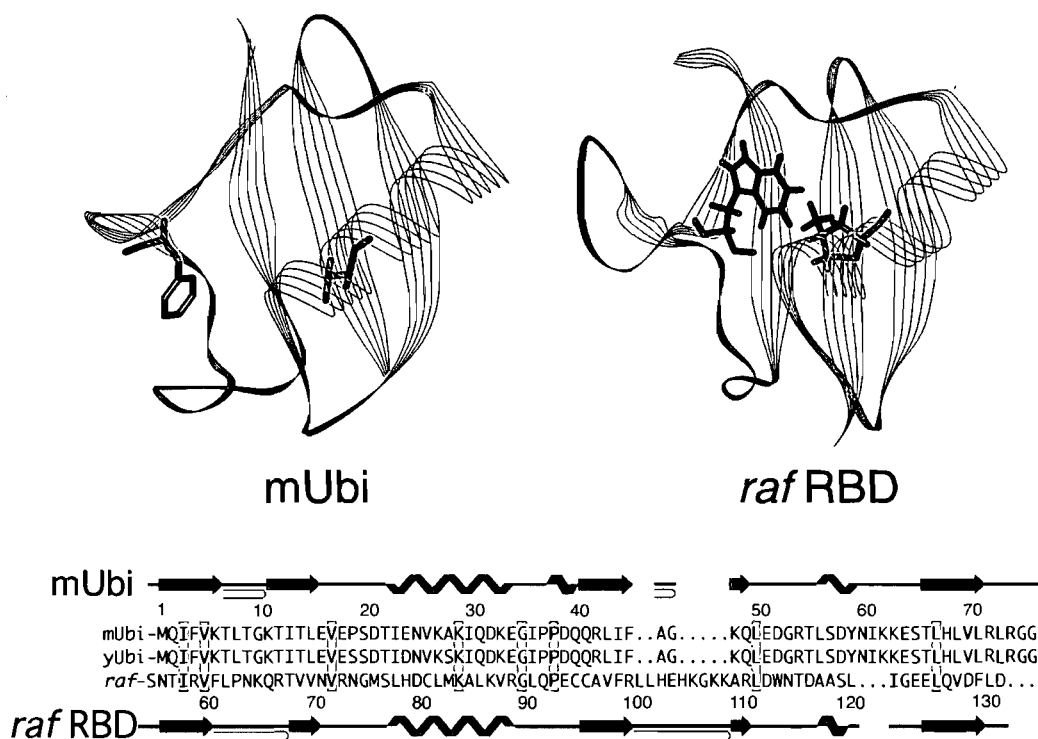


Figure 2-1. Tertiary and secondary ribbon structures of mUbi and *raf* RBD and their primary structure alignment. Amino acids with displayed side chains include Phe 45 and Val 26 for mUbi and Trp 114 and Leu 82 for *raf* RBD. Phe 45, mutated in this study to Trp, and WT Trp 114 of *raf* RBD were used as fluorescent probes to monitor the folding-unfolding transitions, while Val 26 and Leu 82 correspond to the structurally equivalent central hydrophobic residues that were mutated in previous studies and this study. The alignment of the primary structures was derived from both tertiary and secondary structure alignments.³¹ The tertiary and secondary structure representations were based on solution structures of the proteins using WebLab ViewerLite version 3.2 and PDBsum,⁴⁸ respectively (PDB entries 1UBI and 1RFA).

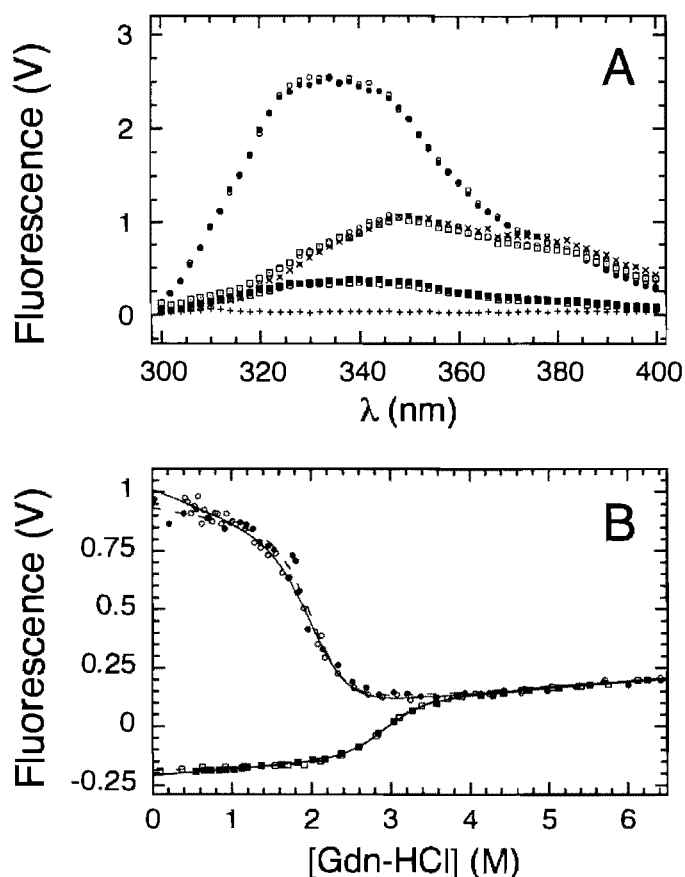


Figure 2-2. (A) Fluorescence emission spectra of yUbi^{F45W} (squares) and *raf* RBD (circles) in native (empty symbols, in 0.46 M Gdn-HCl), renatured (filled symbols, in 0.46 M Gdn-HCl), and denatured states (dotted symbols, in 6.0 M Gdn-HCl). Fluorescence emission spectra of free tryptophan (x) and a 6.0 M Gdn-HCl solution (+) are also shown. All sample concentrations were adjusted to 10 M. (B) Equilibrium stability curves of native (empty symbols) and renatured (filled symbols) states of yUbi^{F45W} (squares) and *raf* RBD (circles) at 25 °C. Fluorescence intensities were taken from the equilibrium end points of unfolding (native) and folding (renatured) kinetic traces (10 s) using 320 and 305 nm cutoff filters for yUbi^{F45W} and *raf* RBD, respectively. The solid lines are results of least-squares-fit analysis of the experimental data based on a two-state folding-unfolding model; see eq 1 in Materials and Methods. The cooperativity values (m) and the $\Delta G^{\text{H}_2\text{O}}$ are reported in Table 2-1. Raw fluorescence data were transformed to set *raf* RBD and yUbi^{F45W} denatured-state fluorescence to the same value.

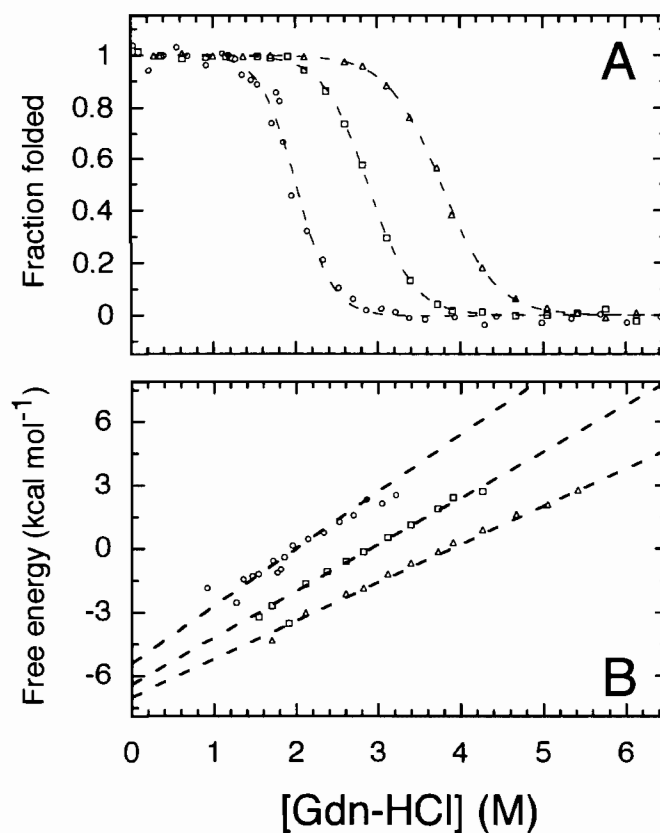


Figure 2-3. Fraction folded vs denaturant (A) and free energy of unfolding (B) of mUbi^{F45W} (triangles), yUbi^{F45W} (squares), and raf RBD (circles) at 25 °C. Fractional equilibrium values and free energies of unfolding were obtained by transforming equilibrium stability curves of unfolding experiments (see Figure 2-2B and Materials and Methods) with a two-state transition model; see eq 2. Thermodynamic parameters are reported in Table 2-1.

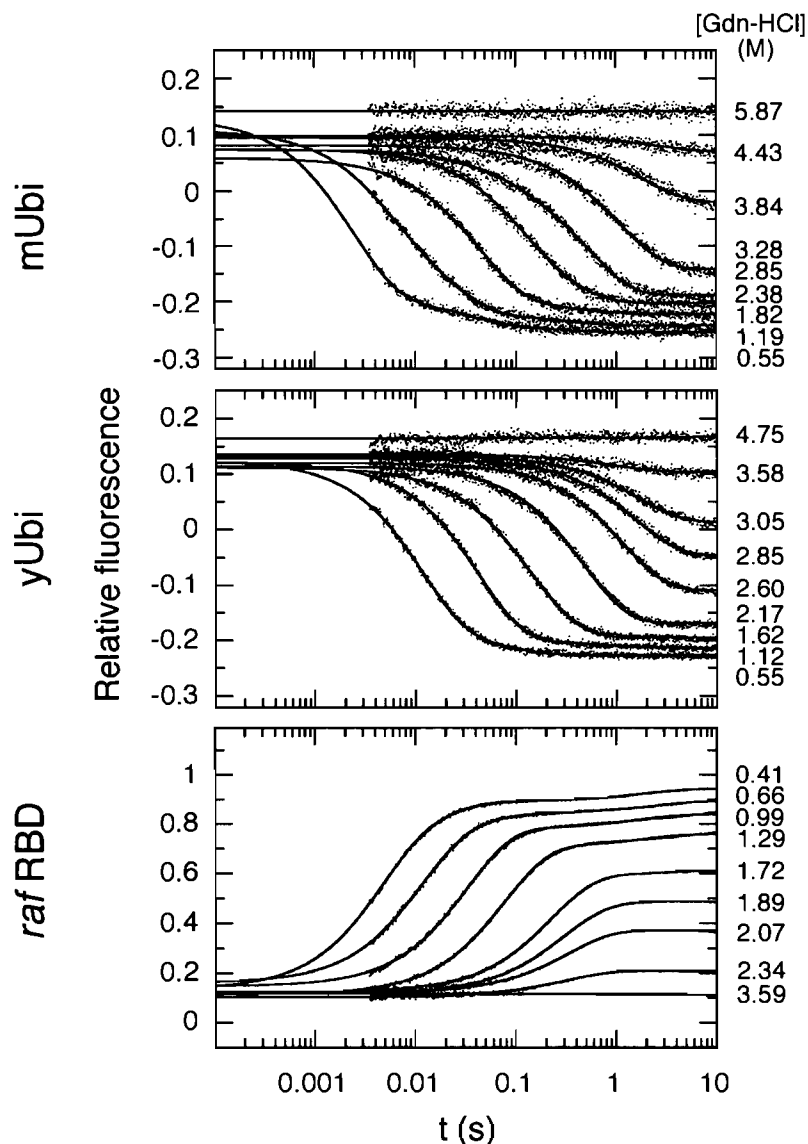


Figure 2-4. Stopped-flow kinetic refolding traces of mUbi^{F45W}, yUbi^{F45W}, and raf RBD at different denaturant concentrations. Denatured protein samples were mixed with various Gdn-HCl solutions using a 1:10 ratio. Fluorescence emission was monitored by using 320 nm (ubiquitin proteins) and 305 nm (raf RBD) cutoff filters. Kinetic traces were fit to single-exponential or multiexponential functions by using nonlinear regression analysis. The noise is reduced at the end of the raf RBD folding experiment by acquiring data in the oversampling mode (fewer acquisitions can be averaged for each point when using the logarithmic distribution of the points on the time axis).

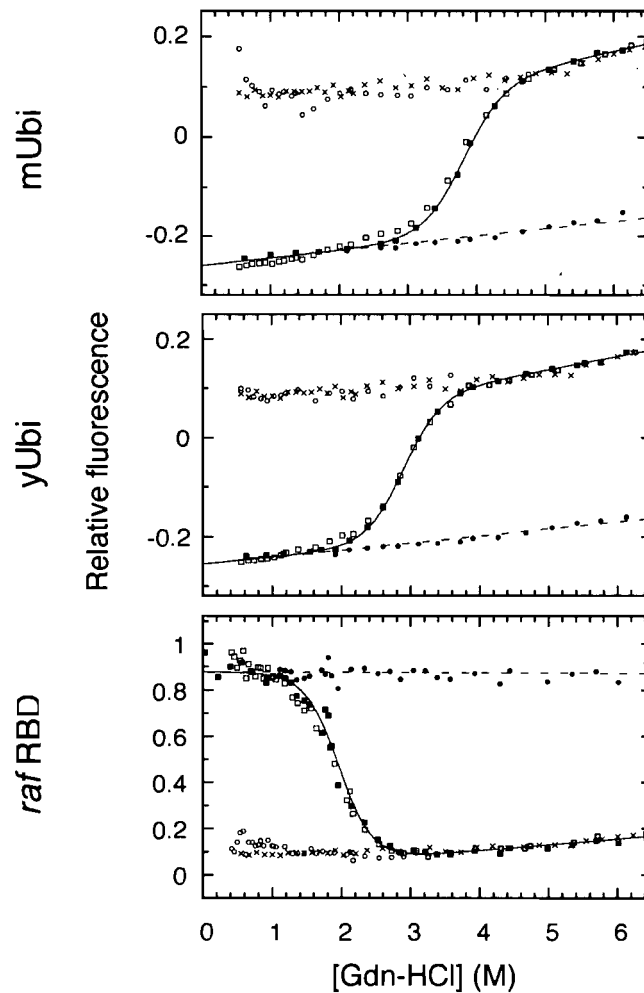


Figure 2-5. Extrapolated initial (circles) and final (squares) fluorescence intensity observed during folding (empty symbols) and unfolding (filled symbols) kinetic experiments with $mUbi^{F45W}$, $yUbi^{F45W}$, and *raf* RBD at 25 °C. Times signs show the fluorescence intensity of free tryptophan vs Gdn-HCl concentration. Extrapolated data were obtained from the simplest exponential functions that fit to the folding and unfolding traces (see Figure 4 and Materials and Methods). Extrapolated infinite points of unfolding experiments (filled squares) were fit to an equilibrium stability curve (eq 1). Fluorescence data obtained from the folding experiments and the free tryptophan samples were set relative to the data from unfolding experiments.

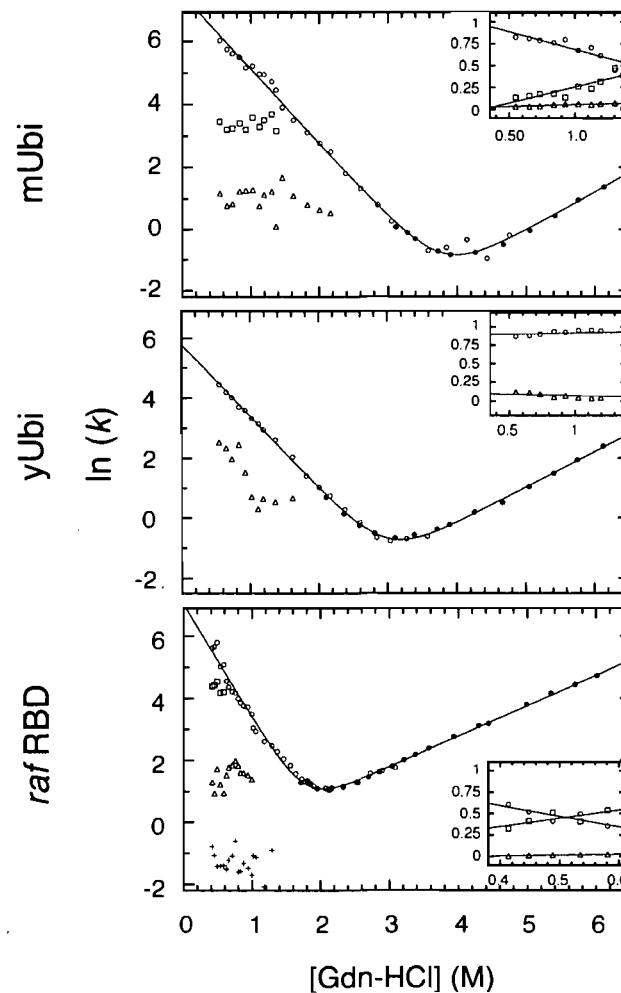


Figure 2-6. Kinetic relaxation rates of folding (empty symbols) and unfolding (filled circle) transitions of mUbi^{F45W}, yUbi^{F45W}, and raf RBD as obtained from the best fit functions of the 10 s stopped-flow kinetic traces (see Figure 2-4). Relative amplitudes of the folding transitions are also shown (inset) with the exception of the slowest phase for raf RBD, which displays ~8% of the total refolding amplitude. Circles, squares, triangles, and times signs correspond to the different observed transitions. The main transition rates of both folding and unfolding experiments were fit to a chevron equation (eq 3).

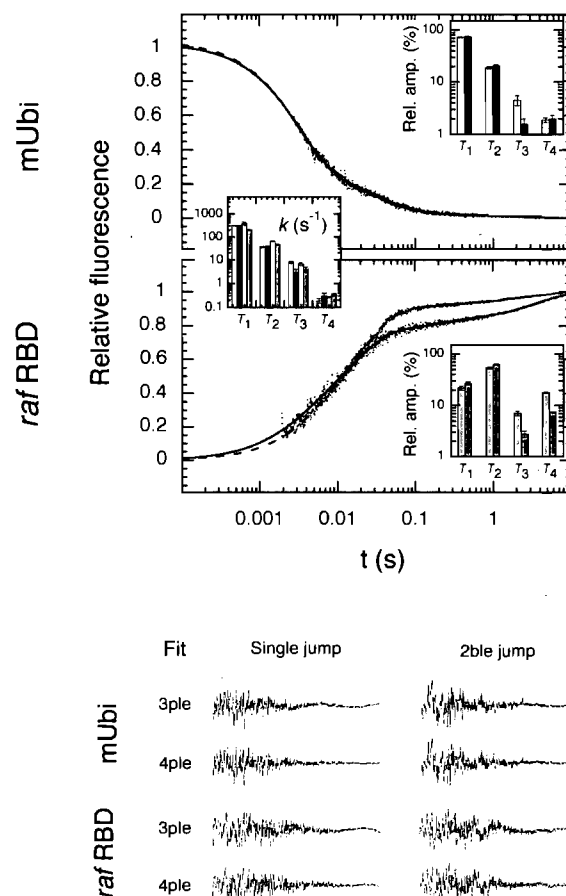


Figure 2-7. Effect of double jump (2ble-jump, rapid unfolding and refolding) on the refolding heterogeneity of His-mUbi^{F45W} (at 0.55 M Gdn-HCl) and His-*raf* RBD (at 0.82 M Gdn-HCl) at 30 °C. Acquisition of the relaxation traces in the oversampling mode revealed an additional fourth transition (T₄) in the refolding of mUbi that was not detected in experiments represented in Figure 2-4 (see residuals of the best three- and four-exponential component fits at the bottom, 3ple and 4ple, respectively). Dashed lines (double-jump experiments) and solid lines (classic single-jump experiments) represent four exponential functions obtained from the best fits of the refolding traces. Rates (k) and amplitudes of the four transitions derived from single-jump experiments (white and light gray rectangles for mUbi and *raf* RBD, respectively) and double-jump experiments (black and dark gray, respectively) are shown as insets. Gdn-HCl concentrations and the temperature were set to obtain the best kinetic resolution of the different transitions.

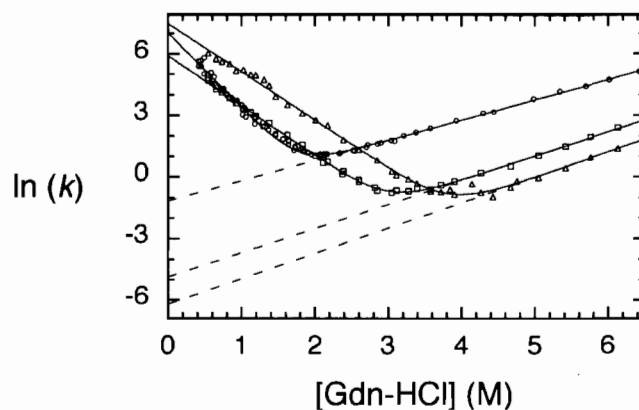


Figure 2-8. Comparative chevron curve plots of $mUbi^{F45W}$ (triangles), $yUbi^{F45W}$ (squares), and *raf* RBD (circles) at 25 C. For $mUbi^{F45W}$, $k_F^{H_2O} = 1800 \text{ s}^{-1}$ and $k_U^{H_2O} = 0.0021 \text{ s}^{-1}$. For $yUbi^{F45W}$, $k_F^{H_2O} = 370 \text{ s}^{-1}$ and $k_U^{H_2O} = 0.0078 \text{ s}^{-1}$. For *raf* RBD, $k_F^{H_2O} = 1100 \text{ s}^{-1}$ and $k_U^{H_2O} = 0.31 \text{ s}^{-1}$. See Table 2-1 for the complete list of the parameters obtained from the fits.

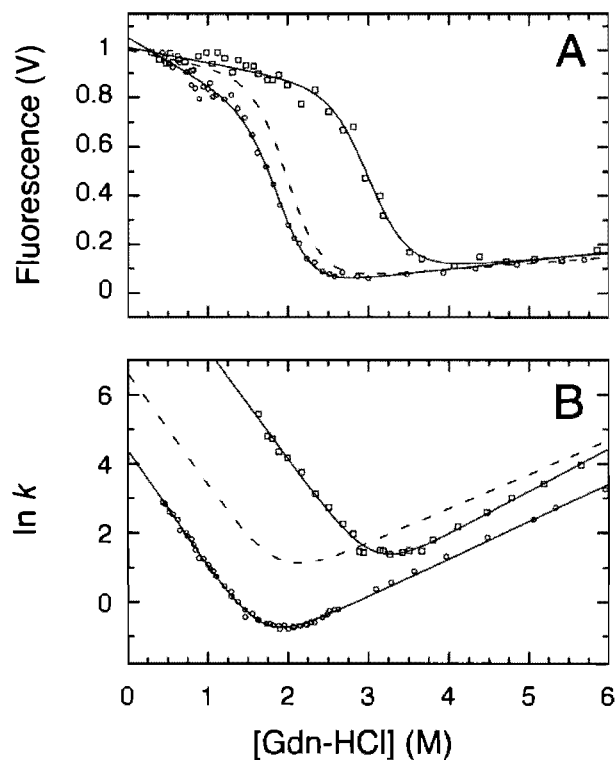


Figure 2-9. Equilibrium stability curves (A) and chevron curves (B) of *raf* RBD at 8 °C (circles) and with 0.4 M Na₂SO₄ (squares). Fluorescence intensities are taken from the equilibrium end point of a 10 s stopped-flow refolding reaction. Dashed lines represent *raf* RBD data at 25 °C.

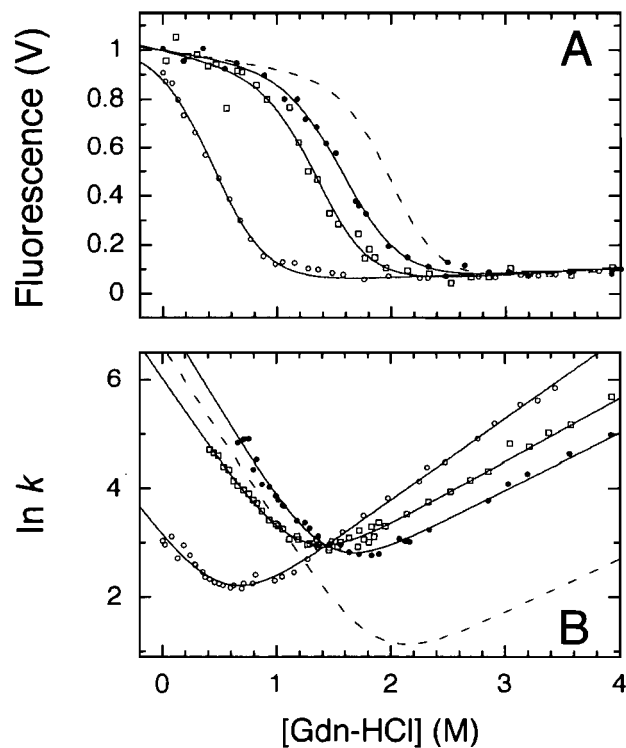


Figure 2-10. Equilibrium stability curves (A) and chevron curves (B) of *raf* RBD mutants L82I (closed circles), L82V (squares), and L82A (circles). All samples were adjusted for an after-mixing concentration of 15 μ M (except for the L82A mutant which was at \sim 5 M for unfolding experiments). Fluorescence intensities are the equilibrium end points of a 10 s stopped-flow unfolding reaction. The dashed line represents the WT *raf* RBD equilibrium and chevron curves at 25 $^{\circ}$ C.

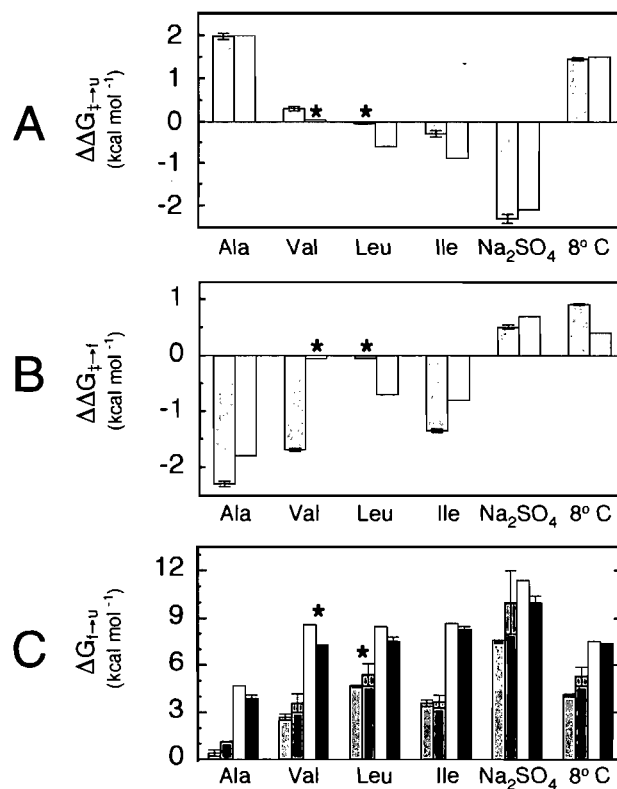


Figure 2-11. Comparison of differences in the free energy barrier of folding (A) and unfolding (B) obtained when various perturbations are used to affect folding-unfolding pathways for both *raf* RBD (gray rectangles) and mUbi (white rectangles; data taken from refs 25 and 26). Mutations to Ala, Val, Leu, and Ile are introduced on a structurally equivalent position in both proteins (see Figure 2-1). The Na₂SO₄ concentration was 0.4 M. (C) Free energies of unfolding extrapolated from the kinetic ΔG values (light gray and white for *raf* RBD and mUbi, respectively) or measured from the equilibrium curves (dark gray and black for *raf* RBD and mUbi, respectively) are compared.

2.8. References

1. Baker, D. (2000) A surprising simplicity to protein folding, *Nature* **405**, 39-42.
2. Baldwin, R. L., and Rose, G. D. (1999) Is protein folding hierarchic? II. Folding intermediates and transition states, *Trends Biochem. Sci.* **24**, 77-83.
3. Baldwin, R. L., and Rose, G. D. (1999) Is protein folding hierarchic? I. Local structure and peptide folding, *Trends Biochem. Sci.* **24**, 26-33.
4. Plaxco, K. W., Simons, K. T., Ruczinski, I., and Baker, D. (2000) Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics, *Biochemistry* **39**, 11177-11183.
5. Dobson, C. M., and Karplus, M. (1999) The fundamentals of protein folding: bringing together theory and experiment, *Curr. Opin. Struct. Biol.* **9**, 92-101.
6. Eaton, W. A., Munoz, V., Hagen, S. J., Jas, G. S., Lapidus, L. J., Henry, E. R., and Hofrichter, J. (2000) Fast kinetics and mechanisms in protein folding, *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327-359.
7. Fersht, A. R. (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, Freeman, New York.
8. Jackson, S. E. (1998) How do small single-domain proteins fold? *Folding Des.* **3**, R81-R91.
9. Myers, J. K., and Oas, T. G. (2002) Mechanism of fast protein folding, *Annu. Rev. Biochem.* **71**, 783-815.
10. Plaxco, K. W., Simons, K. T., and Baker, D. (1998) Contact order, transition state placement and the refolding rates of single domain proteins, *J. Mol. Biol.* **277**, 985-994.

11. Perl, D., Welker, C., Schindler, T., Schroder, K., Marahiel, M. A., Jaenicke, R., and Schmid, F. X. (1998) Conservation of rapid two-state folding in mesophilic, thermophilic and hyperthermophilic cold shock proteins, *Nat. Struct. Biol.* **5**, 229-235.
12. Kim, D. E., Gu, H., and Baker, D. (1998) The sequences of small proteins are not extensively optimized for rapid folding by natural selection, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4982-4986.
13. Riddle, D. S., Santiago, J. V., Bray-Hall, S. T., Doshi, N., Grantcharova, V. P., Yi, Q., and Baker, D. (1997) Functional rapidly folding proteins from simplified amino acid sequences, *Nat. Struct. Biol.* **4**, 805-809.
14. Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., and Dobson, C. M. (1999) Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding, *Nat. Struct. Biol.* **6**, 1005-1009.
15. Martinez, J. C., and Serrano, L. (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved, *Nat. Struct. Biol.* **6**, 1010-1016.
16. Riddle, D. S., Grantcharova, V. P., Santiago, J. V., Alm, E., Ruczinski, I., and Baker, D. (1999) Experiment and theory highlight role of native state topology in SH3 folding, *Nat. Struct. Biol.* **6**, 1016-1024.
17. Alm, E., and Baker, D. (1999) Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11305-11310.
18. Galzitskaya, O. V., and Finkelstein, A. V. (1999) A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11299-11304.

19. Munoz, V., and Eaton, W. A. (1999) A simple model for calculating the kinetics of protein folding from three-dimensional structures, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11311-11316.
20. Larson, S. M., Ruczinski, I., Davidson, A. R., Baker, D., and Plaxco, K. W. (2002) Residues participating in the protein folding nucleus do not exhibit preferential evolutionary conservation, *J. Mol. Biol.* **316**, 225-233.
21. Ferguson, N., Capaldi, A. P., James, R., Kleanthous, C., and Radford, S. E. (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9, *J. Mol. Biol.* **286**, 1597-1608.
22. Kragelund, B. B., Hojrup, P., Jensen, M. S., Schjerling, C. K., Juul, E., Knudsen, J., and Poulsen, F. M. (1996) Fast and one-step folding of closely and distantly related homologous proteins of a four-helix bundle family, *J. Mol. Biol.* **256**, 187-200.
23. Mayer, R. J., Landon, M., and Layfield, R. (1998) Ubiquitin superfolds: intrinsic and attachable regulators of cellular activities? *Folding Des.* **3**, R97-R99.
24. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.* **247**, 536-540.
25. Khorasanizadeh, S., Peters, I. D., Butt, T. R., and Roder, H. (1993) Folding and stability of a tryptophan-containing mutant of ubiquitin, *Biochemistry* **32**, 7054-7063.
26. Khorasanizadeh, S., Peters, I. D., and Roder, H. (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues, *Nat. Struct. Biol.* **3**, 193-205.
27. Roder, H., and Colon, W. (1997) Kinetic role of early intermediates in protein folding, *Curr. Opin. Struct. Biol.* **7**, 15-28.

28. Krantz, B. A., and Sosnick, T. R. (2000) Distinguishing between two-state and three-state models for ubiquitin folding, *Biochemistry* **39**, 11696-11701.
29. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W., and Sosnick, T. R. (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding, *J. Mol. Biol.* **324**, 359-371.
30. Emerson, S. D., Madison, V. S., Palermo, R. E., Waugh, D. S., Scheffler, J. E., Tsao, K. L., Kiefer, S. E., Liu, S. P., and Fry, D. C. (1995) Solution structure of the Ras-binding domain of c-Raf-1 and identification of its Ras interaction surface, *Biochemistry* **34**, 6911-6918.
31. Holm, L., and Sander, C. (1996) Mapping the protein universe, *Science* **273**, 595-603.
32. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues, *Biochemistry* **14**, 4953-4963.
33. Johnsson, N., and Varshavsky, A. (1994) Ubiquitin-assisted dissection of protein transport across membranes, *EMBO J.* **13**, 2686-2698.
34. Picard, V., Ersdal-Badju, E., Lu, A., and Bock, S. C. (1994) A rapid and efficient one-tube PCR-based mutagenesis technique using Pfu DNA polymerase, *Nucleic Acids Res.* **22**, 2587-2591.
35. Scheffler, J. E., Waugh, D. S., Bekesi, E., Kiefer, S. E., LoSardo, J. E., Neri, A., Prinzo, K. M., Tsao, K. L., Wegrzynski, B., Emerson, S. D., et al. (1994) Characterization of a 78-residue fragment of c-Raf-1 that comprises a minimal binding domain for the interaction with Ras-GTP, *J. Biol. Chem.* **269**, 22340-22346.
36. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci.* **4**, 2411-2423.

37. Peterman, B. F. (1979) Measurement of the dead time of a fluorescence stopped-flow instrument, *Anal. Biochem.* **93**, 442-444.
38. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability, *Biochemistry* **38**, 8138-8149.
39. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G., and Kiefhaber, T. (2001) Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics, *Nat. Struct. Biol.* **8**, 452-458.
40. Northey, J. G., Di Nardo, A. A., and Davidson, A. R. (2002) Hydrophobic core packing in the SH3 domain folding transition state, *Nat. Struct. Biol.* **9**, 126-130.
41. Briggs, M. S., and Roder, H. (1992) Early hydrogen-bonding events in the folding reaction of ubiquitin, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2017-2021.
42. Mirny, L., and Shakhnovich, E. (2001) Evolutionary conservation of the folding nucleus, *J. Mol. Biol.* **308**, 123-129.
43. Nauli, S., Kuhlman, B., and Baker, D. (2001) Computer-based redesign of a protein folding pathway, *Nat. Struct. Biol.* **8**, 602-605.
44. Kuhlman, B., O'Neill, J. W., Kim, D. E., Zhang, K. Y., and Baker, D. (2002) Accurate computer-based design of a new backbone conformation in the second turn of protein L, *J. Mol. Biol.* **315**, 471-477.
45. Alm, E., Morozov, A. V., Kortemme, T., and Baker, D. (2002) Simple physical models connect theory and experiment in protein folding kinetics, *J. Mol. Biol.* **322**, 463-476.

46. Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999) Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway, *Struct. Folding Des.* **7**, 1145-1153.
47. Guerois, R., and Serrano, L. (2000) The SH3-fold family: experimental evidence and prediction of variations in the folding pathways, *J. Mol. Biol.* **304**, 967-982.
48. Laskowski, R. A. (2001) PDBsum: summaries and analyses of PDB structures, *Nucleic Acids Res.* **29**, 221-222.

Chapitre 3. Article #2

Suivant la rédaction du premier article, nous avons voulu caractériser les phases additionnelles observées lors du repliement d'ubiquitine (phases qui semblent également être présentes lors du repliement de protéines possédant la même structure). Dans cet article, nous avons donc appliqué l'approche par sondage au tryptophane pour caractériser la nature des phases additionnelles observées lors du repliement d'ubiquitin (voir section 1.2.1.1.). Nous avons ainsi pu déterminer que la première phase détectée lors du repliement de ubiquitine, T_1 (voir Figure 1-11C), représente en fait la formation d'un intermédiaire tardif tandis que la deuxième phase détectée, T_2 , représente la conversion de cet intermédiaire en structure native. Avec l'aide des nouvelles sondes fluorescence très sensibles, nous avons également été en mesure de caractériser l'énergie et le niveau de compaction de cet intermédiaire.

Titre : Multiple tryptophan probes reveal that Ubiquitin folds via a late misfolded intermediate.

Auteurs : Alexis Vallée-Bélisle & Stephen W. Michnick

Publié dans: *Journal of Molecular Biology* (2007) **374**, 791-805

Abbreviations

Ub, mammalian Ubiquitin; Trp, tryptophan; GdnHCl, guanidine hydrochloride; T, transition; I, intermediate; N, native; U, unfolded; A, amplitude; k , rate constants; F, fluorescence; ANS, 8-anilino-1-naphtalenesulfonate; CD, circular dichroism; τ , time constant;

3.1. Abstract

Much of our understanding of protein folding mechanisms is derived from experiments using intrinsic fluorescence of natural or genetically inserted tryptophan (Trp) residues to monitor protein refolding and site-directed mutagenesis to determine the energetic role of amino acids in stabilizing the native- (N), intermediate- (I) or transition- (T) states. However, this strategy has limited use to study complex folding reactions because a single fluorescence probe may not detect and distinguish all low-energy folding intermediates. To overcome this limitation, we suggest that protein refolding should be monitored with different Trp probes, located at various solvent-exposed positions. Here, we demonstrate the utility of this approach by investigating the controversial folding mechanism of Ubiquitin (Ub) using five different single Trp probes located at residue positions 1, 28, 45, 57, and 66. We first show that these solvent-exposed Trp are structurally sensitive and minimally perturbing fluorescent probes for monitoring folding/unfolding of the protein. Using a conventional stopped-flow instrument, we show that ANS and Trp fluorescence detect two distinct transitions during the refolding of all five Trp mutants at low denaturant concentration: T_1 , a denaturant-dependent transition and T_2 , a slower transition, largely denaturant-independent. Surprisingly, some Trp mutants (Ub^{M1W}, Ub^{S57W}) display Trp fluorescence changes during T_1 that are distinct from the expected U→N transition suggesting that the denaturant-dependent refolding transition of Ub is not a U→N transition but represents rather the formation of a structurally distinct I-state (U→I). This U→I transition could be clearly distinguished by combining two Trp mutations Ub^{F45W-T66W} for which the two Trp probes display fluorescence changes of opposite sign during T_1 and T_2 . Global fitting of the folding/unfolding kinetic parameters and additional folding-unfolding

double-jump experiments performed on Ub^{M1W}, a probe with enhanced fluorescence in the I-state compared to U- and N-states, demonstrate that the I-state is stable, compact, misfolded, and on-pathway. These results illustrate how transient low-energy I-states can be efficiently detected and characterized in complex refolding reactions using multiple Trp probes.

3.2. Introduction

In the last decade, the general determinants of protein folding rate and mechanism have started to emerge from the comparison of numerous experimental studies on small proteins that fold without the accumulation of intermediate (I)-states.¹⁻⁵ In contrast, studies on proteins that fold through one or more populated I-states have frequently resulted in controversies⁶ due mainly to the low structural resolution of the probes that are commonly used to monitor fast folding reactions. For example, neither tryptophan (Trp) or ANS fluorescence spectroscopies, nor circular dichroism (CD) or absorbance spectroscopies, can be used to directly assess whether additional folding transitions detected during protein refolding are due to non-obligatory transitions that often accompany "two-state" folding proteins (for example: proline cis-trapped unfolded molecules,⁷ non-prolyl cis-trapped unfolded molecules,⁸ small fraction slower folding proteins⁹⁻¹²), or are related to the presence of sequential I-states in a three-state mechanism.¹³⁻¹⁴ A case in point is long standing debates on the mechanisms of folding of the model protein Ubiquitin (Ub), which have clearly highlighted this limitation over the years.¹⁵⁻²¹ For example, various analyses of Ub refolding using the fluorescence of the genetically inserted Trp 45 (Ub^{F45W}) to probe structure formation led to the proposals that this protein could fold either *via* an early hydrophobic intermediate,^{15, 16, 21, 22} an apparent two-state mechanism,^{17, 18, 23, 24} parallel folding pathways,²¹ or through formation of a some protein aggregates¹⁹ (Figure 3-1). These ambiguities are not likely resolvable by using other spectroscopic techniques since structural changes in small proteins do not often provide sufficient CD or absorbance signal changes to allow for efficient detection, and the ANS probe, despite its high sensitivity, may alter protein folding kinetics.^{14, 25} Alternative approaches using NMR spectroscopy to provide site-specific information on I-states have also been developed and exploited with great success. On the other hand, equilibrium Deuterium/Hydrogen-exchange experiments^{26, 27} cannot determined whether the I-state is on-pathway, pulse Deuterium/Hydrogen-exchange experiments^{28, 29} are in practice quite difficult to perform quantitatively, and

relaxation dispersion NMR experiments³⁰ necessitates that at least 0.5% of I-state population is observable at equilibrium. Our ability to characterize complex refolding reactions would therefore be greatly improved if we could: 1) efficiently detect all the transient low-energy I-states that accumulate during protein folding; and 2) characterize the structural nature of the different transitions that are detected during a refolding kinetic experiment.

Here, we show that we can meet these criteria by monitoring protein refolding using the intrinsic fluorescence of several Trp mutants in which the inserted Trp substitute residues whose side chains are solvent-exposed on the surface of the protein. Previous examples in the literature of Trp substitution as probes of folding or unfolding were done by conservative mutations of partially or completely buried aromatic residues to Trp or *vice versa*.³¹⁻³⁶ The strategy described here is more flexible in that the Trp substitution can be made anywhere on the surface allowing therefore to probe structural transitions in any region of the protein with high efficiency, with minimal or no apparent perturbation of refolding pathways. Specifically, we use different Trp probes in order to discriminate between folding transitions that represent conformational changes that occur in the entire protein population *versus* those that correspond to parallel folding pathways (Figure 3-1). For example, the relative folding amplitude of two transitions representing parallel pathways should remain relatively constant regardless of the probe used since the amplitude of these transitions reflects mostly the fraction of molecules going through each path (Figure 3-1: two-state mechanism with parallel pathways).¹⁴ On the other hand, one should expect that some Trp probes manifest refolding amplitudes that are distinct from the U→N fluorescence change for transitions that represent the formation and/or rearrangement of a structurally distinct I-state (Figure 3-1: three-state mechanism with sequential I-state. Trp fluorescence should display different sensitivity to T_1 and T_2 depending on whether it is located at position A or B in the protein).

We apply our strategy to analyze the controversial refolding mechanism of the small $\beta\beta\alpha\beta\beta\alpha\beta$ mammalian Ub (Figure 3-2 (a)), a 76 amino acid protein that contains no natural Trp, no disulfide bonds and possesses three prolines in the *trans* conformation in the N-state (see recent review by Jackson)³⁷. Recently, Krantz et al.¹⁷ demonstrated that the rate constant of the main refolding transition of Ub^{F45W}, T₁, displays the characteristics of an apparent two-state transition in the absence of a burst phase. Since then, this denaturant-dependent transition has been extensively characterized using both Φ - and Ψ -value analysis.^{23, 24, 39, 40} However, Searle and colleagues recently demonstrated that two additional minor slower folding transitions, T₂ and T₃, are also detected in a proline-less mutant form of Ub, suggesting that these transitions can not be attributed to rate-limiting *cis-trans* prolyl isomerization events (Figure 3-1).²¹ Rapid unfolding-folding experiments performed by us¹⁸ and Crespo et al.²¹ suggest that T₃ represents a small population of slow-folding proteins likely attributable to non-prolyl *cis*-trapped polypeptide⁸ but the nature of T₂ remains unknown. In the present work, Ub refolding was monitored using five different Trp probes in order to characterize both T₁ and T₂, the first and second folding transitions detected when a conventional stopped-flow instrument is used to trigger refolding. Surprisingly, we found that some single Trp mutants displayed refolding amplitude during T₁ that are distinct from the expected U→N fluorescence. This suggests that T₁, the denaturant-dependent folding transition, leads Ub to a structurally distinct I-state that only rearranges in the N-state through T₂, a second rate-limiting step. Global fitting analysis of the rate constants and amplitudes of T₁ and T₂, and additional folding-unfolding double-jump experiments performed with the help of an I-state-sensitive Trp probe allowed us to determine the stability, level of compaction and on-pathway nature of the late I-state. In addition, three out of the five Trp probes studied displayed a significant burst phase during Ub refolding at low denaturant concentration, providing therefore further evidences for accumulation of early species within the dead time of our stopped-flow instrument (< 3.2 ms).²⁰ The approach described here demonstrates how several solvent-exposed Trp probes can provide unambiguous evidence to characterize the different

transitions detected during protein refolding with minimal perturbation of the protein folding landscape.

3.3. Results

The folding landscape of Ub is unaffected by Trp substitutions at solvent-exposed positions

We have created single Trp mutants at different residues whose side-chains are solvent-exposed in the native structure, in order to probe both the local and global structures of Ub with minimal perturbation. In the present study, amino acids Met 1 (1st hairpin), Ala 28 (α -helix), Phe 45 (original Trp location used to study the folding of Ub)¹⁵, Ser 57 (_{3/10}-helix) and Thr 66 (5th β -strand) were chosen in order to monitor the formation of different secondary structure elements of the protein (Figure 3-2(a)). In strong denaturing conditions (6.0 M GdnHCl), the fluorescence spectra of all single Trp mutants were found to be similar to an equimolar solution of N-Acetyl-Tryptophanamide (Figure 3-2(b) -insert). In contrast, different spectra were obtained under native conditions (Figure 3-2(b)) suggesting that Trp located at surface-exposed locations can provide sensitive probes to monitor the folding/unfolding transition. Equilibrium-unfolding curves of each mutant were then monitored in order to determine that the stability of Ub is not altered by Trp insertion. All five curves (Figure 3-2(c)) fitted well to a two-state model showing single transitions of similar steepness (similar *m*-value; see Table 3-1). The extrapolated free energies of unfolding in water ($\Delta G^{\text{H}_2\text{O}}$) were also found to be similar to the values obtained for *wild type* Ub using NMR or CD spectroscopies to monitor unfolding (31 and 28.0 kJ mol⁻¹, respectively).¹⁵ Mutants Ub^{T66W} (27.0 kJ mol⁻¹) and Ub^{A28W} (35.2 kJ mol⁻¹) displayed the lowest and highest stabilities, respectively.

We then assessed the unfolding kinetics of the different mutants using GdnHCl-jump experiments (Figure 3-3). All kinetic traces fitted well to a single exponential function (Figure 3-3 (a), left) and the amplitude of the transitions accounted well for the N→U fluorescence change expected at various denaturant concentrations (Figure 3-3 (a), right). Unfolding rate constants of the different mutants fitted well *versus* denaturant concentration to a two-state chevron curve model and were found to be similar within one order of magnitude (Figure 3 (b)). Both the stabilities and *m*-value estimated from the chevron curve were found to be similar to values obtained in equilibrium experiments (Table 3-1). This suggests that no folding I-state is significantly populated during folding/unfolding of Ub above 2 M GdnHCl. The different mutants also displayed similar unfolding rate denaturant dependencies. Their estimated β_T -values (Table 3-1), which provide a measure for the relative compactness of the transitions-state, were similar to that of Ub^{F45W} (0.68) [values near zero or one indicate a level of compaction similar to the U-state or the N-state, respectively].³ These results suggest that above 2 M GdnHCl, the folding/unfolding transition-states of all mutants have similar compactness and that Trp insertions at solvent exposed locations on the protein do not significantly perturb the folding/unfolding mechanism. To ensure that the different Trp insertions do not alter the folding landscape of Ub at low denaturant concentration, the folding kinetics of all mutants were also studied using ANS fluorescence as a probe (fluorescence quantum yield of this dye markedly increases when bound to exposed clusters of hydrophobic residues in partially folded proteins).⁴¹ All mutants showed a comparable biphasic decrease of fluorescence during refolding (Figure 3-4) and the rate constants of both transitions detected, kf_1 and kf_2 , were found to be similar to the rate constants of T₁ and T₂ (Table 3-2). Taken together, these results suggest that Trp insertion at solvent-exposed positions has little impact on either the stability or the folding/unfolding landscape of Ub and provide sensitive and structure-specific probes to monitor its refolding.

The denaturant-dependent folding transition leads to a structurally distinct late I-state

We then assessed how the different Trp probes report Ub folding transitions (Figure 3-5). At low denaturant concentration, two exponential terms were sufficient to model the recovery of native-like fluorescence of all mutants between 3.2 ms and 300 ms of refolding time with the exception of Ub^{T66W}, which required only a single exponential (Figure 3-5 (a-b); see Material and Methods for fitting procedure). The two folding transitions displayed rate constants similar to those obtained with ANS fluorescence: at 0.45 M GdnHCl, kf_1 varied from 121 s⁻¹ to 327 s⁻¹ depending on the mutant (average value obtained with ANS: 134 s⁻¹), and kf_2 varied from 11.5 s⁻¹ to 17.5 s⁻¹ (average value obtained with ANS: 19.9 s⁻¹) (Table 3-2). At low denaturant concentration, kf_1 fitted well to the expected two-state folding transition (Figure 3-5 (c): open circles fit to black dotted line), and the denaturant dependency of kf_2 was found to be relatively independent of the GdnHCl concentration (open squares). The only folding transition detected with Trp 66 displayed a rate constant and denaturant dependency similar to the second transition (13.9 s⁻¹ in 0.45 M GdnHCl). Signs of apparent burst phase amplitude were also detected at low denaturant concentration for Ub^{M1W}, Ub^{A28W} and Ub^{T66W} since a discrepancy was observed between the fluorescence signal extrapolated to the initiation time of the folding reaction, F_0 , and the linear extrapolation of the U-state fluorescence (dotted line; Figure 3-5 (b)). This suggests that an additional folding transition could also take place during the dead time of the stopped-flow instrument^{20, 42}. Concerning the refolding amplitudes of the two detected transitions (A_1 and A_2), both were found to vary strongly depending on probe location. The majority of the expected refolding amplitude occurred during T_1 for Ub^{A28W} and Ub^{F45W} while most or all of the expected refolding amplitude was detected during T_2 for Ub^{S57W} and Ub^{T66W}, respectively (Table 3-2 and Figure 3-5 (b)). Surprisingly, A_1 was opposite to that for the expected U→N transition for Ub^{S57W}. In contrast, the fluorescence signal of Ub^{M1W} increased during T_1 to a level that is twice the value of its N-state signal (Figure 3-5 (b)): F_1

versus F_2) before decreasing back to native-like level during T_2 . These refolding profiles obtained at a protein concentration of 15 μM at pH 7.0 were similar at different protein concentration (down to 1 μM) and at lower pH (pH 5.0) (see Supplementary information). In summary, all Trp probes detected two transitions in Ub refolding (with the exception of Trp 66, see below “Protein kinetic I-state highlighted using a combination of Trp probes”), and T_1 and T_2 showed similar rates but very different relative amplitudes depending on the Trp position (Table 3-2).

Knowing that the different Trp mutations do not significantly perturb the Ub folding landscape, we then compared the refolding amplitudes of T_1 and T_2 obtained with different Trp probes in order to assess the nature of these transitions (Figure 3-1). To begin with, the analysis of Ub^{S57W} demonstrates that T_1 cannot represent a direct path from U to N since A_1 is inverted compared to the expected U→N fluorescence amplitude change (A_{U-N}) in absence of an apparent burst phase (Table 3-2). This suggests that the denaturant-dependent folding transition of Ub, does not lead the protein directly to the N-state but rather to a structurally distinct late I-state (U→I). The term “late” is used in order to distinguish the I-state formed during T_1 from the state that may be formed earlier within the dead time of the stopped-flow instrument. The stability of this late I-state can be further inferred from the denaturant effect on both F_1 , the fluorescence level obtained after T_1 , and kf_1 , the observed rate constant for the I-state formation (Figure 3-5 (b-c): Ub^{S57W}). Above 1.5 M GdnHCl, F_1 decreases in a cooperative manner while kf_1 displays an upward deviation. This suggests that the late I-state becomes less stable than the U-state above 1.5 M GdnHCl and therefore no longer accumulates [kf_1 increases being attributable to the enhancement of the I-state unfolding rate].

Evidence that T_2 represents the rate limiting rearrangement of the late I-state into the N-state (I→N) is provided by the correlation observed between kf_2 roll over at low denaturant concentration and the amount of I-state that accumulates after T_1 . Specifically,

as the late I-state becomes destabilized above 1.5 M GdnHCl, kf_2 displays the characteristic of an apparent two-state transition (Figure 3-5 (b-c): Ub^{S57W}). Similar observations are also obtained from the analysis of Ub^{M1W}, with the exception that Trp 1 cannot distinguish T₁ from T₂ above the I-state mid-point of denaturation because their amplitudes (A_1 and A_2) become of similar sign [the formation of less than half of the Ub^{M1W} I-state population displays a fluorescence amplitude that is smaller than the expected U→N transition] and kf_1 and kf_2 are too similar to be distinguished (within 3 fold). This limitation in probe efficiency becomes even more apparent and problematic in the case of Trp 28 and Trp 45 because these probes cannot even distinguish T₁ from T₂ above 1 M GdnHCl since their amplitudes change in the same direction and rate constants are too similar. Interestingly, Trp 66 represents an extreme case of probe inefficiency since it is not sensitive to the formation of the late I-state (T₁). This suggests that Trp 66 displays similar levels of fluorescence both in the U- and I-states U→I. Finally, no evidence of correlation is observed between the burst phase and the other transitions: kf_1 does not deviate from linearity as the burst phase is enhanced in Ub^{M1W} or in Ub^{A28W}, and A_2 remains similar in Ub^{M1W} as the burst phase disappears around 1.5 M GdnHCl. These results add further evidence against the hypothesis that T₂ represents a parallel pathway in which a fraction of the protein slowly rearranges into the N-state after having been trapped in a slow folding I-state (burst phase) (see Figure 3-1). In the absence of any direct kinetic measurements, and without any evidence of kinetic coupling with either T₁ or T₂, the burst phase remains therefore a subject for further research (see Discussion).

Protein kinetic I-state highlighted using a combination of Trp probes

As described in the previous section, only two of the five Trp probes studied (Trp 1 and 57) were sensitive to the formation and rearrangement of the late I-state during Ub refolding. Trp 28 and 45 only displayed small minor fluorescence changes during the I-state

rearrangement into the N-state (T_2) and Trp 66 did not detect the formation of the I-state (T_1). However, two Trp mutants that display opposite fluorescence changes during T_1 and T_2 can also be combined in the same mutant in order to confirm that the observed transitions represent the formation and rearrangement of a structurally distinct intermediate and not parallel folding pathways. For example, Trp 45 fluorescence is quenched during the $U \rightarrow I$ transition while Trp 66 fluorescence is enhanced during the $I \rightarrow N$ transition (Figure 3-6 (a), black dots). Therefore, the fluorescence of double Trp mutant $Ub^{F45W-T66W}$ should provide an ideal probe to highlight the late I-state since its fluorescence intensity is expected to decrease during the I-state formation (T_1), and increase during its rearrangement into the N-state (T_2) (see Figure 3-6 (a), dotted line). We investigated the refolding of $Ub^{F45W-T66W}$ versus denaturant concentration (Figure 3-6 (b)) and found that the rates and amplitudes of both folding transitions in the double Trp mutant were in close agreement with the values expected from combining Trp 45 and Trp 66 in the same protein (Figure 3-6 (a) compare red filled circles with black dotted line). In addition, adding Trp 66 to Ub^{F45W} , or Trp 45 to Ub^{T66W} also allowed us to clearly highlight the late I-state in the refolding of both Ub^{F45W} , and Ub^{T66W} since Trp 66 alone could not detect the formation of the I-state and Trp 45 alone could hardly distinguish the late I-state from the N-state.

Global fitting analysis shows that the late I-state is compact, misfolded, and on-pathway

Having found that the denaturant-dependent transition does not represent a direct path from the U- to N-states but rather the formation of a structurally distinct late I-state, we needed to confirm whether this alternative state is on- or off-pathway. To do so we simultaneously fitted kf_1 , kf_2 , k_U and F_1 , for mutants Ub^{M1W} , Ub^{S57W} , and $Ub^{F45W-T66W}$ to both on- and off-pathway models over all denaturant concentrations (see Material and Methods for fitting procedure). The similarity between the rates for formation (kf_1) and rearrangement (kf_2) of the I-state allowed us to distinguish between both mechanisms.⁴³

The kinetic data of all three mutants fitted well to an on-pathway model (red lines in Figures 3-5 (b-c) and 6 (c)) and the best parameters obtained for the fits allowed for an estimate of the stabilities and the m -value of the different mutants (Table 3-3). On the other hand, the kinetic data obtained for Ub^{M1W} and Ub^{F45W-T66W} could not be fitted to an off-pathway mechanism because the I-state unfolding rate (k_{IU}) could not be found equal to or higher than the I-state rearrangement rate into the N-state (k_{IN}) at low denaturant concentration (data not shown; see also Figure 3-7 (b)).

The denaturant dependencies of kf_1 (m_{UI}), F_1 (m_I), and kf_2 (m_{IN}) obtained from the best fit also suggest that all three mutants display similar folding mechanism (Table 3-3). From the m_{UI} -values, we can estimate that the relative degree of compaction of the transition-state for the formation of the late I-state, β_{TS1} , lies between 0.53 and 0.59, depending on the Trp position. From the m_I -values, one can also estimate that the compactness of the I-state, β_I , lies between 0.82 and 0.90. The stability of this I-state is also found to be relatively high: its average value being around half of the N-state stability (15.5 kJ mol⁻¹), although lower in the case of Ub^{S57W} (10.1 kJ mol⁻¹). Finally, the rate constants for the I→N transition of each mutant (kf_2) were all found to slightly increase with the denaturant concentration (positive m_{IN} -values) suggesting that the second T-state is less compact than the I-state (β_{TS2} : 0.66 – 0.79). This suggests that the late I-state rearrangement proceeds through a partial unfolding event since on average, approximately 13 % of the buried surface in the I-state must be re-exposed to the solvent before folding can proceed to the N-state. This necessary local unfolding event further suggests that some regions of the Ub structure may be misfolded in the I-state.

Double jump experiments on mutant Ub^{M1W} confirm that the I-state is on-pathway

To further characterize the late misfolded I-state, we took advantage of the fluorescence characteristics of Trp 1 that displays a much higher fluorescence signal in the

I-state than in either the U- or N-states. Rapid folding/unfolding double jump experiments were performed on Ub^{M1W} in order to directly detect the late I-state unfolding transition (I→U) and therefore confirm the accuracy of the parameters obtained from the global fit analysis (Figure 3-7). After 101 ms of refolding time in 1.59 M GdnHCl (no burst phase observed at this GdnHCl concentration), a rapid GdnHCl jump to 3.76 M showed two distinct Ub populations unfolding with rate constants of 45 s⁻¹ and 0.30 s⁻¹ (Figure 3-7 (a): red dots). In contrast, after 261 ms of refolding, only the most stable protein population with an unfolding rate constant similar to the N-state (0.35 s⁻¹), remained detectable. The unfolding rate constant of the late I-state, k_{IU} , and its dependence on denaturant concentration were therefore directly determined using rapid unfolding jump experiments at various denaturant concentration after 101 ms of refolding time in 1.59 M GdnHCl (Figure 3-7 (a) insert). Both k_{IU} and m_{IU} were found to be identical to the values predicted by the global fitting analysis (Figure 3-7 (b): filled circles fit to the red dotted line). Moreover, the on-pathway nature of the I-state can now be directly established since the linear extrapolation of k_{IU} in water (red dotted line) is estimated to be around one order of magnitude slower than k_{IN} , the rate constant for the rearrangement of I-state into the N-state (1.6 s⁻¹ *versus* 10.3 s⁻¹, respectively). In addition, the folding/unfolding double jump experiment also demonstrates that T₁ and T₂ represent the sequential formation of two different species during Ub refolding (Figure 3-7 (c)). First, the amplitude of the fastest unfolding transitions (circles), corresponding to the I-state population, could be well modeled *versus* refolding time by using kf_1 and kf_2 , the rate constant for the formation and rearrangement of the late I-state. Unfortunately, the rate of appearance of the late I-state could not be directly assessed due to the large dead time associated with the double mixing experiments (101 ms). Also, the rate of formation of the N-state (squares: amplitude of the slowest unfolding transition *versus* refolding time) was similar to that of the disappearance of the I-state and was accurately modeled using an on-pathway three-state model using the same parameters.³ Additional evidence in support of the sequential mechanism is also suggested by the fact that this modeling also correctly accounts for all N-state formation

(N-state population is extrapolated to zero at the initiation of refolding) implying therefore that the entire N-state population originates from the rearrangement of the late I-state.

3.4. Discussion

One of the main advances provided by using several Trp probes to study protein refolding is that it allows to discriminate between transitions that represent conformational changes that occur in the entire protein population *versus* those that correspond to parallel folding pathways (Figure 3-1). In the present study, the utilization of Trp 1 and Trp 57 allowed us to demonstrate for the first time that T_1 and T_2 do not represent parallel pathways, as this has been recently suggest by Crespo et al.,²¹ but rather the formation and rearrangement of a late sequential I-state in a three-states mechanism (Figure 3-5). In addition, an appropriate combination of Trp probes (Trp 45 and Trp 66) allowed to confirm the presence of this structurally distinct states (Figure 3-6). The advantage of selecting Trp probes that display large fluorescence changes to study a specific transition was also illustrated in the double jump experiments when using the large fluorescence variation of Trp 1 to detect the I→U transition (Figure 3-7). This specific probe allowed us to determine the stability, degree of compaction and *on-pathway* character of the I-state. Finally, the utilization of different single Trp probes also provided new indirect evidences suggesting that an early I-state may also accumulate during the deadtime; Ub^{M1W}, Ub^{A28W} and Ub^{T66W} all displayed an apparent burst phase amplitude at low denaturant concentration (Figure 3-5 (b)) (see below: “*Ub may also refold through an early I-state*”). In summary, our results describe how complex kinetic refolding reactions can be efficiently analyzed, first by using several Trp probes to detect and identify all relevant folding I-states, and then by characterizing each I- or T-states using double-jump experiments (Figure 3-7). Further insights into the nature of I- and T-states could be provided by Φ -value³ or Ψ -value⁴⁴

analysis on specific Trp mutants that display the highest fluorescence sensitivity to a state or transition of interest. We believe that such an integrated strategy will provide a versatile approach to unambiguously characterize non-two-state protein folding mechanisms.

Folding mechanism of Ub

During the last decade, several groups have published more than twenty folding kinetic studies on Ub (see review by Jackson)³⁷ including both its complete Φ ²⁴ and Ψ -value analysis.²³ In most of these works, Trp 45 was the only fluorescence probe used to monitor Ub refolding and T₁, the major refolding transition detected, was considered to be the only significant folding transition.^{17-19, 23, 24, 45} In this study, we present direct kinetic evidence to suggest that this transition represents instead the formation of a misfolded I-state displaying around half of the stability of the N-state and up to 85% of its degree of compaction. A new mechanism for Ub folding based on these results is presented in the next section.

Ub folds via a late misfolded I-state. The energy and relative compactness of the different states and transition-states detected during the refolding of Ub^{M1W} (Table 3-3) are illustrated in Figure 3-8 at various GdnHCl concentrations. In the absence of denaturant, the energy barrier between U and TS₁, ΔG_{U-TS1} , is found to be significantly smaller than the energy barrier between the late I-state (I_L) and TS₂, ΔG_{I-TS2} , allowing therefore the late I-state to accumulate. On the other hand, as the GdnHCl concentration is raised, the U-state gets more stable relative to the other more compact states and ΔG_{U-TS1} and ΔG_{I-TS2} become closer in height; at 2 M GdnHCl both the U- and I-states, and TS₁ and TS₂ display similar energies. Above this denaturant concentration, the Ub folding mechanism displays an apparent two-state behaviour since the I→N rearrangement is no longer rate limiting. However, the late I-state could not be detected in conventional unfolding experiments since

ΔG_{I-TS1} , the unfolding energy barrier for the I-state is smaller than ΔG_{N-TS2} , the energy needed to unfold the N-state, over all GdnHCl concentrations (Figure 3-8). Still, one of the most remarkable features of Ub late folding I-state is its high degree of compaction, even closer to the N-state than that of the last transition-state (β_I and β_{TS2} being 0.84 and 0.79, respectively, for Ub^{M1W}). This suggests that some buried residues in the late folding I-state have to be re-exposed to solvent so that the polypeptide chain can reach the last transition state. But what is the nature of this rearrangement, or what is causing such a large late rate barrier in Ub folding that is not found in other small simple “two-state” folding proteins? A plausible hypothesis could be that some specific non-native interactions may prevent the misfolded I-state from rapidly accessing the N-state (specific energetic frustration)^{46, 47}. In this case, ΔG_{I-TS2} would represent the energy that is needed to disrupt these interactions. Another possibility could be that the complexity of the Ub fold requires the stepwise formation of a specific I-state in order to allow the protein to fold efficiently (topological frustration)^{48, 49}. In this case, the misfolded I-state would be attributable to the complexity of the Ub fold rather than to specific non-obligatory detail of the sequence. Evidence from the literature would seem to support the latter hypothesis as described below.

The most suggestive evidences concerning the nature of T_2 were likely reported in the first kinetic folding study on *wild type* Ub using NMR and pulse Deuterium/Hydrogen exchange experiments to monitor amide H-bond formation *versus* refolding time.⁵⁰ In this pioneering work, Briggs & Roder found that amide H-bonds of residues 59, 61, and 69 were protected at a slower rate (similar to kf_2) than the majority of the other residues as Ub refolds in 1 M GdnHCl (25 °C, pH 5.0). This observation provided evidence for the subsequent stabilization or packing of the C-terminal segment (3/10 helice and last β -brand).⁵⁰ This result is also consistent with the refolding amplitude obtained in the present study since Trp 57 and Trp 66, both located in this specific region of the protein, displayed the highest fluorescence sensitivity to T_2 .

Additional evidence in support of this model is also found in a recent study by Searle *et al.*, which provide interesting data on a mutant in which the last C-terminal β -strand is destabilized.²¹ In this study, the authors found that the mutant Ub^{T66A-H68A} (two mutations at solvent-exposed residue positions) produced a ~ 20 -fold acceleration in the N-state unfolding rate (from around 0.007 s^{-1} to 0.14 s^{-1} in water), which revealed another unfolding transition with a rate constant that is very similar to the unfolding rate constant of our late I-state (0.27 s^{-1} in water *versus* 0.44 s^{-1} in water for Ub^{F45W-T66W} k_{IU}). Interestingly, the energy diagram presented in Figure 3-8 predicts that the unfolding transition of the Ub late I-state could be detected in single-jump unfolding experiments if a mutation would render ΔG_{N-TS2} similar to ΔG_{I-TS1} . The double mutation T66A-H68A appears to have selectively destabilized the N-state but not the late I-state, suggesting that the last C-terminal segment of Ub is not yet found in its native-like β -strand conformation in the late I-state. Φ -value analysis performed on Ub is also in agreement with this model even though no direct signs of late folding rearrangement was detected using fluorescence of Trp 45 on any of the single point mutants studied.²⁴ Indeed, unfolding kinetic data of all the single point mutants indicate that the native-like interactions in the C-terminal part of the protein are not yet formed in TS₂.^{24, 40} In contrast, Ψ -value analysis performed by Sosnick *et al.* suggests that native-like interactions between the last β -strand of Ub and its adjacent β -strands might already be formed in the transition-state.^{23, 40} On the other hand, all Ψ -value experiments in this work have been conducted at a relatively high denaturant concentration where the rearrangement is no longer rate limiting and the I-state no longer accumulates (around 2.0 M Gdn-HCl).

In summary, evidence presented above suggests that the late folding rearrangement of Ub may implicate the subsequent stabilization or packing of the C-terminal segment of the protein ($_{3/10}$ -helice and last β -brand). On the other hand, these results do not indicate why this late rearrangement necessitates such a high rate barrier. Concerning this question, it is interesting to note that a late I-state has also been detected during the refolding of GB1,

a protein with similar topology to Ub but no amino acid identity.⁵¹ The degree of compaction of both the I-state and the last transition-state were also found to be similar to that of Ub (β_I and $\beta_{TS2} = 0.85$). Equally, *raf* RBD, a Ub superfold member, was also found to fold *via* two transitions displaying similar rate constants and rate denaturant dependencies to Ub.¹⁸ Φ -value analysis further suggests that the apparent “two-state” folding transition-state of *raf* RBD has a structure that is similar to the transition-state of Ub.⁵² Such similarities in folding landscapes in the absence of sequence detail resemblance suggest that the particular fold complexity of Ub may be at the origin of both the late I-state and transition-state. With specific Trp probes that distinguish between both the U \rightarrow I and I \rightarrow N transitions, Φ - or Ψ -value analysis will efficiently allow the determination of the structure of the misfolded I-state and that of the late rate barrier and whether they represent common features of the folding pathway of Ub superfamily members.

Ub may also refold through an early I-state. The strongest evidence in support of the presence of an earlier folding transition during Ub refolding has been provided recently, by Roder and colleagues, using a rapid mixing device.²⁰ In this study, two additional minor transitions have been detected during the first refolding milliseconds of Ub^{F45W} ($\tau \sim 150 \mu\text{s}$ and 1.8 ms). Further analyses are required to determine whether such kinetic complexities that take place on the 0.1 – 2 ms time scale (within the dead time of our stopped-flow instrument) suggest the formation of early on-pathway intermediates or more or less specific collapse events. In the present study, two results add further evidence for accumulation of early species within the dead time of our stopped-flow instrument (< 3.2 ms). First, the decrease in ANS fluorescence intensity observed as Ub refolds suggests the presence of an early collapse state since a high fluorescence level of this probe is generally observed only upon ANS binding to hydrophobic patches, which are not likely found in completely unfolded proteins.⁴¹ The apparent burst phase observed at low denaturant concentration with Trp probes located at position 1, 28, and 66 provides additional evidence. However, it is important to note that care must be taken when interpreting an

apparent burst phase, since it has been demonstrated that the extrapolation of F_0 (the fluorescence signal at the beginning of the refolding experiment) can be affected when more than 50% of a transition is hidden in the dead time of the mixing instrument (which is the case for T_1 at the lowest GdnHCl concentration).¹⁷ Other studies have also provided evidence supporting both the presence and absence of accumulation of an early I-state during the refolding of Ub. For example, Crespo *et al.* found a consistent discrepancy between the equilibrium and kinetic stabilities and m -values they obtained on many Ub mutants which they attribute to the presence of a rapidly forming I-state (mean values: $m_{eq} = 10.6 \pm 0.6 \text{ kJ mol}^{-1} \text{ M}^{-1}$; $m_{kin} = 8.5 \pm 0.6 \text{ kJ mol}^{-1} \text{ M}^{-1}$). On the other hand, our results do not corroborate this observation since both the three-state and the two-state analyses (above 2 M GdnHCl) of our kinetic data correctly estimated our equilibrium parameters (see Table 3-1 and 3-3). It has also been proposed that some non-two-state behavior detected during Ub refolding at low denaturant concentration could be attributable to aggregation.¹⁹ However, in the present study, refolding curves of Ub^{M1W} -a Trp mutant displaying a high fluorescence sensitivity for all phases- displayed similar kinetic and burst phase amplitude at different protein concentrations (down to 1 μM) or at lower pH (5.0) (see Supporting information). Finally, small-angle X-ray scattering measurements performed by Sosnick and colleagues showed no sign of a collapsed state upon rapid dilution of Ub U-state into 0.75 M GdnHCl at pH 7, 20 °C.⁵³ With the recent advances in rapid mixing devices,⁵⁴ our capacity to perform efficient Φ - or Ψ -value analysis on Ub fastest refolding transitions will be greatly enhanced using the Trp mutants that display the largest fluorescence change (burst phase) during the dead time of our conventional stopped-flow instrument (Ub^{M1W}, Ub^{A28W}, Ub^{T66W}).

General implications of this study

It has long been recognized that obligatory on-pathway intermediates can be present in the refolding mechanism of apparent two-state folding proteins if the barrier for formation of the I-state is higher than that for its rearrangement to the N-state,^{55, 56} or if the I-state free energy is higher than the U- or N-states.⁵⁷ In the present work, we described another case where a three-state system displays the characteristics of a two-state mechanism: when the fluorescence signal and stability of a late low-energy I-state is similar to that of the N-state (mutants Ub^{A28W} and Ub^{F45W}). In addition, we also demonstrated that an apparent roll-over of the refolding rate constant at low denaturant concentration can represent the signature of a late I-state if the probe used is only sensitive to the I→N rearrangement (Trp 66). These results illustrate why protein folding kinetics have to be monitored with multiple Trp probes in order to detect the accumulation of all low-energy I-states and to assess whether heterogeneity in folding traces or roll over of the rate constant at low denaturant concentration represent evidence for late I-states that would have been “missed” due to the limitations of using only a single fluorescent probe. More generally, it has been shown recently that more than 50% of the reported two-state folding proteins, either the *wild type* protein or some variants, show folding or unfolding rate constant deviations which suggest the presence of a structurally similar late folding rate barrier ($\beta_{TS2} \sim 0.9$).⁵⁸ Combined with the fact that many late I-states have been recently reported in the folding of different small proteins (e.g. Ub, Im7,⁵⁹ En-HD,⁶⁰ Barnase,⁶¹ Rd-apocyt b₅₆₂,⁶² the third domain of PDZ,⁶³ cytochrome c¹²), these results suggest that late I-states may well represent a common feature of protein folding.

3.6. Materials and methods

Protein Constructs and Expression. The coding sequence of Mammalian Ubiquitin was fused to a hexa-His tag at its N-terminus –MHHHHHHG. It has been found that a His-tagged variant of Ub is not susceptible to transient aggregation.¹⁹ Trp mutations and expression of the different Ub mutants were carried out as previously described.¹⁸ His-tagged proteins were purified at high concentration (~600 μ M) either in denatured (6M Gdn) or native conditions using Ni-NTA resin (Qiagen) and eluted from the column with 25 mM Acid Acetic.

Data Collection. All experiments were performed in 50 mM sodium phosphate buffer pH 7.0⁶⁴ at 30 ± 0.1 °C in order to allow for proper kinetic resolution between T_1 and T_2 . Fluorescence emission spectra were obtained using a Varian Cary Eclipse spectrophotometer and represent the average of 3 acquisitions (0.1 s sampling time) taken every 2 nm (bandwidth of 2.5 nm) with λ_{ex} = 281 nm. All folding/unfolding data are the average of three experiments acquired using an Applied Photophysics SX18.MV stopped-flow fluorimeter with λ_{ex} = 281 nm (2.5 nm bandwidth) by reading fluorescence intensity using a 320 nm cut-off filter (in ANS experiment: λ_{ex} = 350 nm using a 395 nm cut-off). Equilibrium stability curves were generated using the equilibrium end point data of 10 s folding (Figure 3-2) or unfolding kinetic traces (Figure 3-3). Single-jump experiments were carried out by 1:10 mixing of denatured or native proteins (5.0 M and 1.0 M GdnHCl, respectively) with different GdnHCl concentration. Double jump experiments were carried out by refolding denatured protein (5.0 M GdnHCl) with a six-fold dilution followed by a second mixing (2:5) with different GdnHCl concentration. Final GdnHCl concentration of each experiment was determined by refractive index measurements with an Abbe 60 refractometer (Bellingham & Stanley L^{td}) using an independent set of standard dilutions. Final protein concentrations in single or double jump experiments were ~15 μ M and ~5 μ M,

respectively.⁶⁵ A concentration of 200 μ M of ANS was used in order to monitor the refolding of ~ 15 μ M of Ub molecules.

Data Analysis was performed using the non-linear regression analysis program Kaleidagraph (version 3.6 synergy Software, pCS Inc.). Equilibrium data were fit to a two-state model assuming that the fluorescence intensity of the N and U-states and ΔG_{U-N} are linear functions of Gdn-HCl concentration.³ Proper adjustment of the time axis of single and double mixing kinetic data set was obtained by performing an independent dead-time calibration of the stopped-flow instrument using N-bromosuccinimide quenching of N-Acetyl-Tryptophanamide fluorescence.⁶⁶ 0.7 ms (single mixing) and 2 ms (double mixing) were added to the x-axis and experimental data acquired before 3.2 ms (single mixing) and 3.5 ms (double mixing) were removed. The accurate minimal ageing time of the double jump experiment (101 ms) was determined by comparing refolding traces of Ub^{M1W} obtained on both single and double mixing mode. Unfolding kinetic traces could be well described by a single exponential function in the single jump experiments. Refolding kinetic traces at low denaturant concentration were fitted with two exponential terms between 3.2 ms and 300 ms. We use this time window that focuses on the first two phases and not the whole time course in order to improve fitting procedure. This allowed to capture all N-like fluorescence recovery while getting rid of the additional minor folding transitions that always displayed small amplitude with all Trp probes (T₃ and T₄ –not shown in this publication, see ¹⁸). All the refolding and unfolding traces obtained for each Trp mutants were set relative to the fluorescence of the unfolded-state (between 2 M and 6 M Gdn), which was set to zero in absence of GdnHCl (unless specified (Figure 3-3 (a))). Unfolding rate constants (k_U) obtained in GdnHCl concentrations above 2 M (Figure 3-3 (b)) were fitted to a classic two-state chevron curve ³. Folding kinetic parameters of both the I-and N-states were obtained from the best simultaneous fit of kf_1 , kf_2 , k_U , and F_1 over all denaturant concentration. kf_1 , the observed rate constant for formation of the I-state, was fitted to a chevron curve providing therefore information on k_{UL} , k_{IU} , m_{UL} , m_{IU} (and m_I and

ΔG_{UI} , indirectly). F_1 , the fluorescence obtained after T_1 , was fitted to a standard “two-state” equilibrium curve in order to obtain an estimation of m_1 and ΔG_{UI} . Finally, k_f and k_u , the observed rate for formation and unfolding of the N-state, respectively, were fitted to a three-state on- or off-pathway mechanism in order to obtain k_{IN} , k_{NI} , m_1 and ΔG_{UI} using the equation provided in ⁶⁷. Standard deviations were obtained from the best fit of the data.

ACKNOWLEDGMENTS

The authors acknowledge F.X. Campbell-Valois, Tobin Sosnick and Sheena Radford for helpful discussions, Emily Manderson for critical reading of the manuscript, M. Fyfe for sequencing, and Jeffrey W. Keillor for providing access to the stopped-flow instrument. This work was supported by the NSERC of Canada (Grant 194582-00). SWM holds the Canada Research Chair in Integrative Genomics.

3.7. Tables

Table 3-1. Stabilities and m -values of Ub^{M1W}, Ub^{A28W}, Ub^{F45W}, Ub^{S57W}, and Ub^{T66W} obtained using equilibrium (e) folding (F) or unfolding (U) experiments or kinetic (k) unfolding experiments at 30 °C.

	F/U	$\Delta G^{\text{H}_2\text{O}^a}$	m^b	β_T^c
Ub ^{M1W}	F (e)	33.3 (5.3)	10.0 (1.6)	
	U (e)	26.8 (3.3)	8.3 (1.1)	
	U (k)	31.8 (1.4)	10.0 (0.5)	0.68 (0.05)
Ub ^{A28W}	F (e)	35.2 (4.6)	10.8 (1.4)	
	U (e)	26.8 (1.1)	8.7 (0.4)	
	U (k)	29.1 (0.7)	8.7 (0.2)	0.67 (0.02)
Ub ^{F45W}	F (e)	28.0 (2.0)	8.8 (0.8)	
	U (e)	26.8 (1.1)	8.7 (0.4)	
	U (k)	30.3 (1.5)	10.1 (0.5)	0.68 (0.03)
Ub ^{S57W}	F (e)	30.3 (1.9)	9.0 (0.6)	
	U (e)	30.6 (1.5)	9.5 (0.5)	
	U (k)	30.9 (1.6)	9.8 (0.5)	0.70 (0.03)
Ub ^{T66W}	F (e)	27.0 (6.1)	9.4 (2.1)	
	U (e)	30.6 (1.5)	9.5 (0.5)	
	U (k)	28.6 (4.9)	9.5 (1.7)	0.66 (0.07)

^a The free energies of folding (ΔG) are in kJ mol⁻¹. Kinetic ΔG are determined from the folding and unfolding rate constants extrapolated in water from the chevron curve fits of the kinetic unfolding experiments: $\Delta G_{\text{UI}}: -RT \ln[k_{\text{UN}} / k_{\text{NU}}]$.

^b The m -values are in kJ mol⁻¹ M⁻¹. Kinetic m -values (k) are the sum of the slope of the denaturant dependencies for the unfolding and folding rates ($m_{\text{NU}} - m_{\text{UN}}$) multiplied by the factor RT (gas constant times temperature).

^c β_T -values are determined from the following ratio $1 - m_{\text{NU}} / m_{\text{e(U)}}$.

Table 3-2. Rate constants (k) and amplitudes (A) of the transitions detected during the refolding of Ub^{wt}, Ub^{M1W}, Ub^{A28W}, Ub^{F45W}, Ub^{S57W}, and Ub^{T66W} in 0.45 M GdnHCl using Trp or ANS as fluorescent probes.^a

Probes		kf_1	kf_2	A_0	A_1	A_2	A_{U-N}
ANS	wt	146	21.0	-	0.58	0.42	1
	M1W	114	15.0	-	0.60	0.41	1
	A28W	156	25.0	-	0.57	0.43	1
	F45W	151	20.8	-	0.61	0.39	1
	S57W	128	17.7	-	0.66	0.34	1
	T66W	120	20.9	-	0.60	0.40	1
Trp	wt	-	-	-	-	-	-
	M1W	327	11.5	-0.56	-0.58	0.64	-0.50
	A28W	162	17.5	-0.40	-0.36	-0.08	-0.84
	F45W	178	15.6	0.11	0.75	0.20	1.06
	S57W	121	13.7	0.00	-0.08	0.65	0.57
	T66W	-	13.9	-0.40	-	-0.64	-1.04

^a Rate constant (s⁻¹) and refolding amplitudes of the two folding transitions detected during Ub refolding between 3.2 and 300 ms of refolding time in 0.45 M GdnHCl using ANS or Trp fluorescence as probes. kf_1 , kf_2 , A_1 , and A_2 represents the rate constants and refolding amplitudes of T_1 and T_2 , respectively. A_0 represents the difference in amplitude between the fluorescence signal extrapolated at the initiation time of refolding and the expected unfolded fluorescence signal (burst phase). A_{U-N} represents the refolding amplitude expected for the U→N transitions which was set to 1 in the case of the refolding traces monitored with ANS. Error is smaller than 10% for rate constants and smaller than 0.01 for the refolding amplitudes.

Table 3-3. Folding and unfolding parameters obtained from a three-state on-pathway global fit analysis of k_{f1} , k_{f2} , k_U and F_1 for Ub^{M1W}, Ub^{S57W}, and Ub^{F45W-T66W}.

		Thermodynamic parameters ^a				Kinetic parameters ^b			
Transitions→		U-N (e)	U-N ^c (k)	U-I ^c (k)		U→I	I→U	I→N	N→I
Ub ^{M1W}	ΔG	33.3 ± 5.3	33.6 ± 1.9	16.8 ± 0.9	k	1053	1.6	10.3	0.013
	m	10.0 ± 1.6	11.1 ± 0.8	9.3 ± 0.3	m	-6.7	2.64	0.60	2.38
	β_T^d	-	-	$\beta_I = 0.84$	β_T^d	$\beta_{TS1} =$	0.59	$\beta_{TS2} =$	0.79
Ub ^{S57W}	ΔG	30.3 ± 1.9	30.4 ± 2.1	10.1 ± 0.6	k	284	5.15	9.5	0.003
	m	9.0 ± 0.6	9.3 ± 1.7	8.4 ± 0.7	m	-5.16	3.24	2.25	3.13
	β_T^d	-	-	$\beta_I = 0.90$	β_T^d	$\beta_{TS1} =$	0.56	$\beta_{TS2} =$	0.66
Ub ^{F45W-T66W}	ΔG	24 ± 6	32.9 ± 3.2	19.5 ± 2.1	k	1012	0.44	6.0	0.030
	m	10 ± 2	12.4 ± 1.8	10.2 ± 1.2	m	-6.6	3.55	0.75	3.02
	β_T^d	-	-	$\beta_I = 0.82$	β_T^d	$\beta_{TS1} =$	0.53	$\beta_{TS2} =$	0.76

^a The free energies of unfolding in water (ΔG) are in kJ mol⁻¹. The m -values are in kJ mol⁻¹ M⁻¹. (e) or (k) identify values that were determined from equilibrium or kinetic experiments, respectively.

^b The rate constant (k) are in units of s⁻¹. Kinetic m -values are the slope of the denaturant dependencies for the different transitions multiplied by the factor RT (gas constant times temperature).

^c The thermodynamic parameters determined from the best fits of the kinetic data. ΔG_{UI} : $-RT \ln[k_{UI} / k_{IU}]$; ΔG_{UN} : $-RT \ln[(k_{UI} \times k_{IN}) / (k_{IU} \times k_{NI})]$; $m_I = m_{IU} - m_{UI}$; $m_N = m_{IU} - m_{UI} + m_{NI} - m_{IN}$.

^d β_T -value is the degree of compaction of a state or a transition-state relative to the U-state (0) and N-state (1) determined from the following ratio using m_N obtained from the kinetic experiment: $\beta_I = m_I / m_N$; $\beta_{TS1} = -m_{UI} / m_N$; $\beta_{TS2} = (m_I + m_{UI}) / m_N$; where I, TS1 and TS2 represent the I-state and the transition-states for the formation and rearrangement of the I-state, respectively.

3.8. Figures

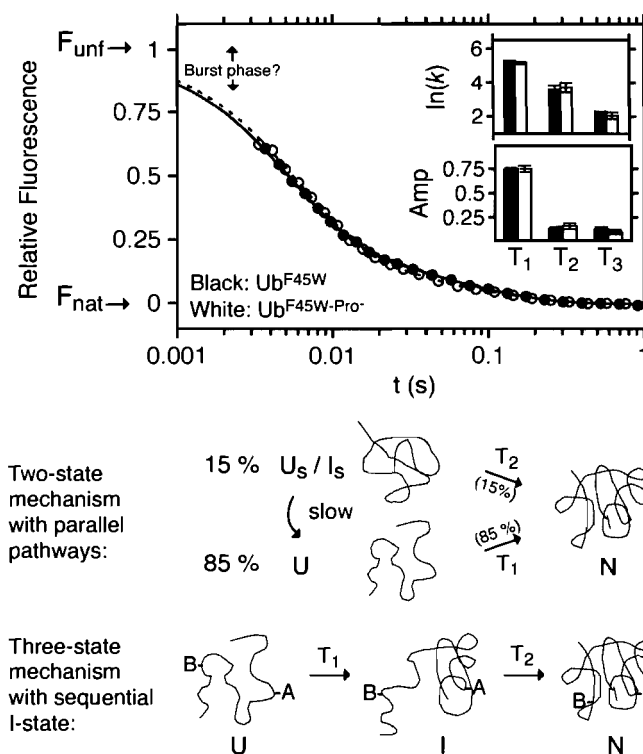


Figure 3-1. Ub refolding probed by Trp 45 fluorescence. Refolding traces of Ub^{F45W} (closed symbols) and its proline-less mutant (open symbols) are better modeled with the sum of three exponential functions between 3.2 ms and 1 s at 0.45 M GdnHCl. Amplitudes and rate constants of the different transitions (T) are shown insert. It has been demonstrated that T_1 displays the characteristics of an apparent two-state transition,¹⁷ and rapid unfolding-folding experiments^{18, 21} suggest that T_3 represents a small population of slow-folding proteins likely attributable to non-prolyl *cis*-trapped polypeptide.⁸ T_2 could either represent a parallel channel involving around 15% of slower folding proteins (U_s or I_s : two-state mechanism with parallel pathways) or the structural rearrangement of a sequential I-state (three-state mechanism with sequential I-state). In this latter case, T_1 and T_2 should display very different refolding amplitudes depending on the probe location on the protein (ex: location A or B). The presence of a faster folding transition, inferred from a burst phase, is also the subject of debate.^{17, 18, 20, 21}

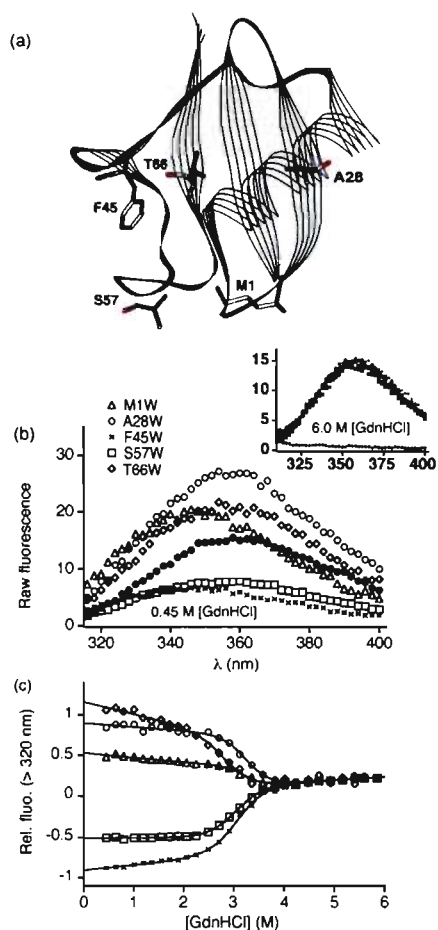


Figure 3-2. Equilibrium characterization of the different single Trp mutants. (a) Ribbon representation of Ub structure. Amino acids with displayed side-chains are mutated to Trp in the present study (pdb code: 1UBI)³⁸. (b) Fluorescence spectra of the Trp mutants in 0.45 M and 6.0 M GdnHCl (insert). Closed circles represent the fluorescence spectrum of an equimolar solution of N-Acetyl-Tryptophanamide (10 μ M). *Wild type* Ub (+) showed no significant intrinsic fluorescence when compared with the other single Trp mutants (insert). (c) Equilibrium stability curves of the different Trp mutants monitored using Trp fluorescence intensity (320 nm cut-off filter). Extrapolated free energies of unfolding in water (ΔG^{H_2O}) and their denaturant dependencies obtained from the best fits are reported in Table 3-1.

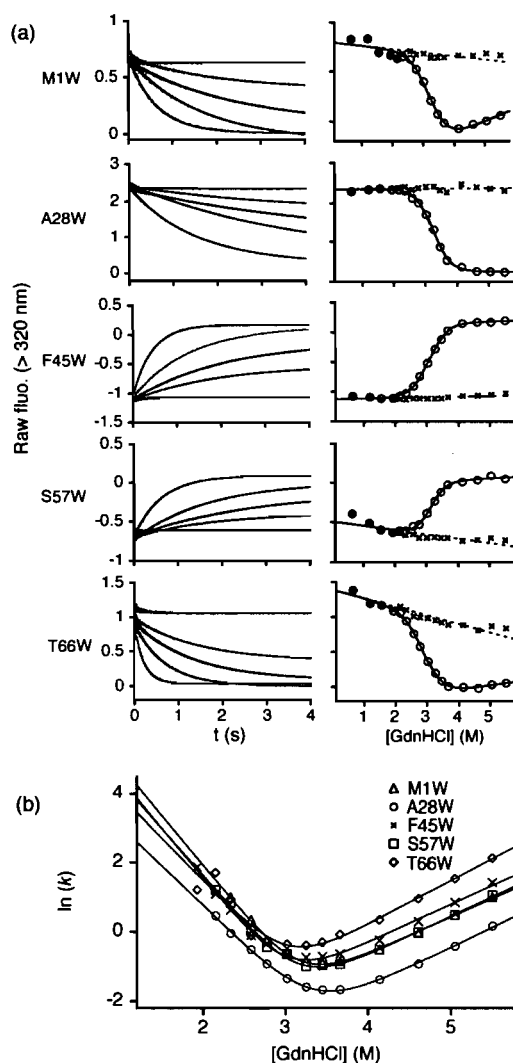


Figure 3-3. Unfolding kinetic characterization of the different single Trp mutants. (a) *left*: Unfolding kinetic traces of the different Trp mutants in 5.05 M (orange), 4.14 M (red), 3.45 M (green), 3.02 (blue) and 2.15 M GdnHCl fitted to single exponential functions; *right*: fluorescence signal extrapolated at the initiation (crosses) and end point (circles) of the kinetic unfolding trace fit well to the expected N-like signal (dotted line) and a two-state equilibrium curve, respectively; (b) unfolding rate constants of the Trp mutants *versus* GdnHCl concentration fit to a two-state mechanism (chevron curve). See Table 3-1 for thermodynamic parameters extrapolated from the best fits.

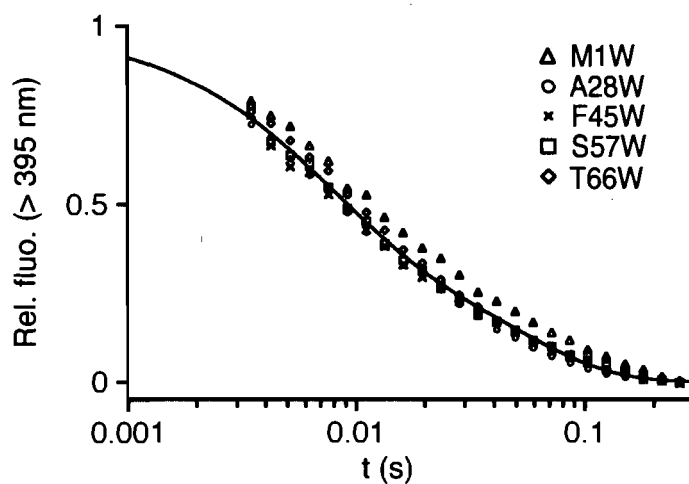


Figure 3-4. Folding kinetic traces of the different Trp mutants monitored using ANS fluorescence (395nm cut-off filter). All traces were well fitted with the sum of two exponential functions. The line represents the fit obtained for *wild type* Ub refolding trace. Rate constants and amplitudes of the two transitions detected are reported in Table 3-2. Refolding traces contain fewer points in order to distinguish all mutants.

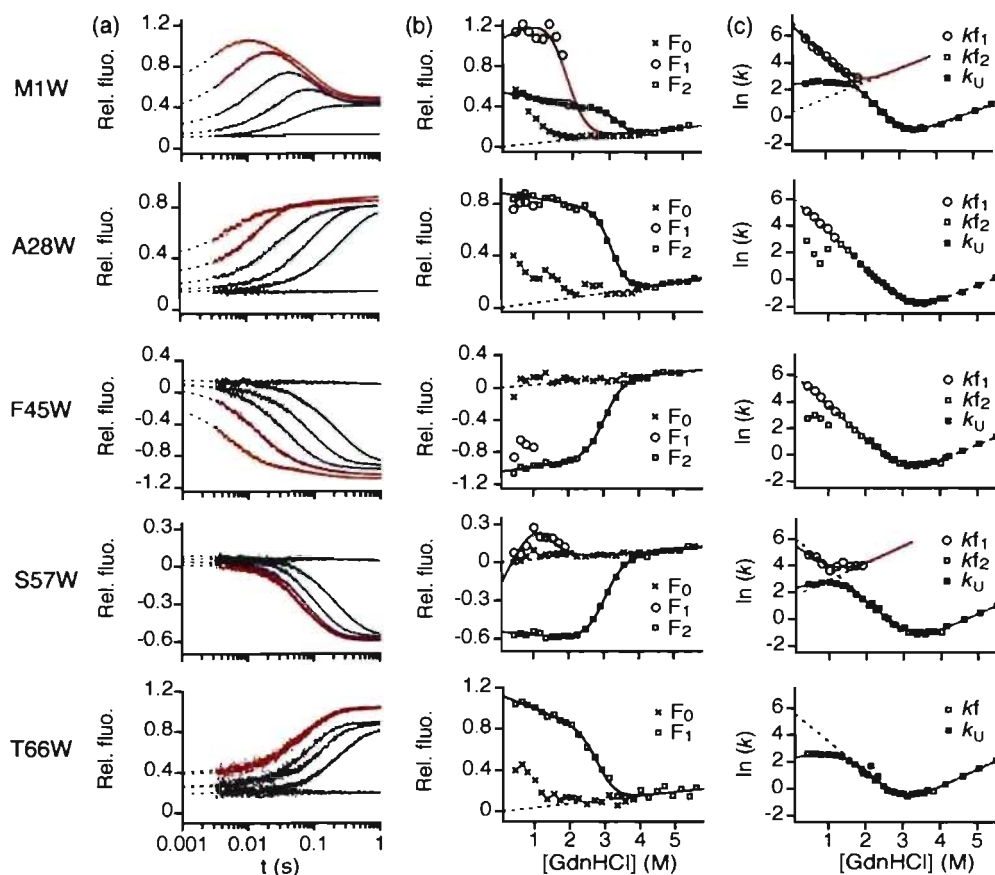


Figure 3-5. Folding kinetic characterization of the different single Trp mutants monitored using Trp fluorescence (320 nm cut-off filter). (a) Refolding traces obtained between 3.2 and 300 ms at various GdnHCl concentrations (orange: 0.45 M; red: 0.80 M; purple: 1.19 M; blue: 1.55 M; green: 1.89 M; black: 4.02 M). (b) Fluorescence values extrapolated at the initiation of refolding (F_0), after T_1 (F_1), and after T_2 (F_2); F_2 fits well to a two-state folding model using the parameters obtained from the equilibrium experiment (black line). Red line corresponds to a two-state fit of F_1 , the fluorescence signal obtain after T_1 . Dashed line corresponds to the linear extrapolation of the expected fluorescence of the U-state. Apparent burst phases can be detected by comparing F_0 with this latter value. (c) Rate constants of folding transitions T_1 and T_2 (k_{f1} and k_{f2}) together with the unfolding rate constant (k_U). Black dashed line corresponds to a two-state chevron curve fit of k_f and k_U under the apparent two-state conditions ($[\text{GdnHCl}] > 2 \text{ M}$). Red lines correspond to the three-state on-pathway fit of the data (see Materials and methods for fitting procedure and Table 3-3 for parameters obtained from the best fits).

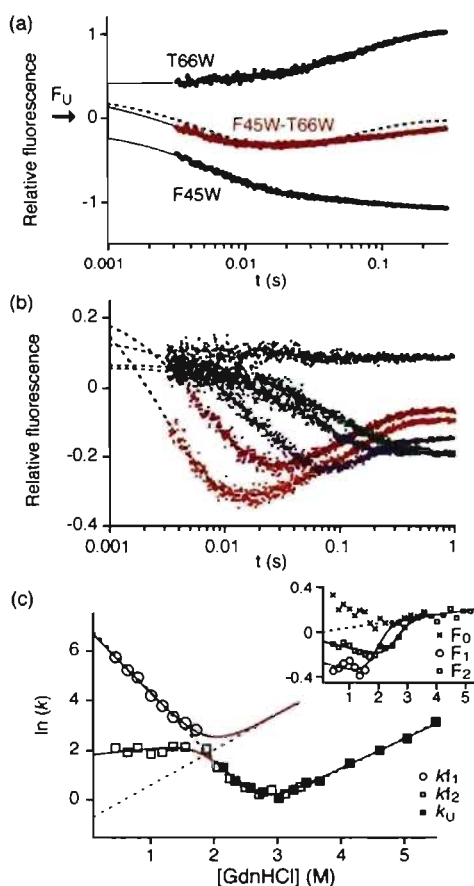


Figure 3-6. Folding kinetic characterization of the double Trp mutant Ub^{F45W-T66W} monitored using Trp fluorescence (320 nm cut-off filter). (a) Refolding traces of single Trp mutants Ub^{F45W}, and Ub^{T66W} (black filled circles), and the predicted (dotted line) and experimentally determined (red filled circle) refolding traces of the double Trp mutant Ub^{F45W-T66W}. (b) Refolding traces of Ub^{F45W-T66W} at various GdnHCl concentrations (orange: 0.45 M; red: 0.80 M; purple: 1.19 M; blue: 1.55 M; green: 1.89 M; black: 4.02 M). (c) Rate constants of the folding and unfolding transitions. Black dashed line corresponds to a two-state chevron curve fit of k_f and k_u in the apparent two-state conditions ($[GdnHCl] > 2$ M). Red lines correspond to a three-state on-pathway fit of the data (see Materials and methods for fitting procedure and Table 3-3 for parameters obtained from the best fit). Insert: fluorescence values extrapolated at the initiation of refolding (F_0), after T_1 (F_1), and after T_2 (F_2). F_2 fits well to a two-state unfolding model using the parameters obtained from the equilibrium experiment (black line). Red line corresponds to a two-state fit of F_1 , the fluorescence signal obtain after T_1 . Dashed line corresponds to the linear extrapolation of the expected fluorescence of the U-state.

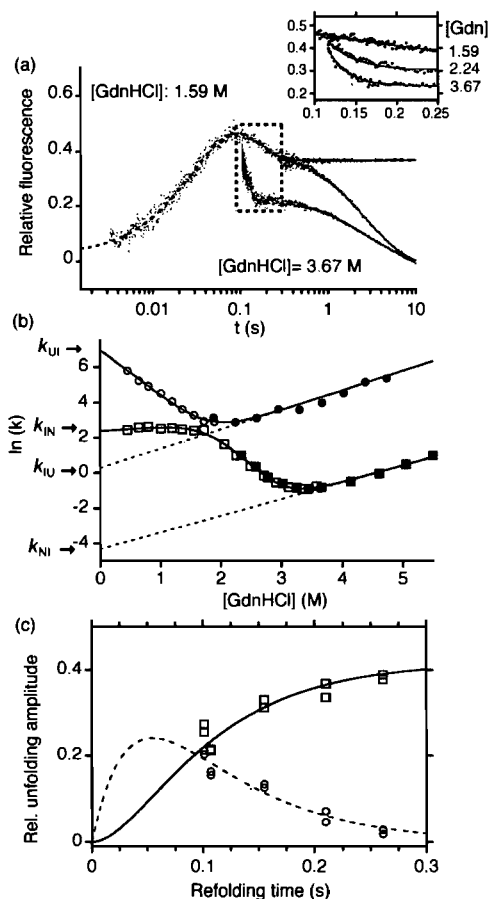


Figure 3-7. Characterization of Ub late I-state using folding/unfolding double-jump experiments on Ub^{M1W}. (a) Kinetic unfolding traces (red dots) obtained following a rapid GdnHCl jump (from 1.59 M to 3.76 M) after 101 ms (two unfolding transitions: 45 s⁻¹ and 0.30 s⁻¹) and 261 ms of refolding (one unfolding transition: 0.35 s⁻¹—similar to the N-state unfolding rate constant). The GdnHCl concentration used to refold Ub was selected in order to minimize burst phase and allow sufficient kinetic resolution between T₁ and T₂. Insert: unfolding the I-state at different GdnHCl concentration—magnification of (a). (b) Three-state modeling of the rate constants obtained from the folding/unfolding kinetic experiments on Ub^{M1W} (fit taken from Figure 3-5 (c)). The unfolding rate constant of the late I-state obtained from the double jump experiment (closed circles) fit perfectly with the predicted k_{IU} (red dotted line). (c) Kinetics of appearance or disappearance of both the I- (circles) and N-states (squares) *versus* folding time fit to an on-pathway sequential mechanism³ with rate constant for formation and rearrangement of the I-state set to kf_1 and kf_2 (25 s⁻¹ and 13 s⁻¹, respectively; $F_1 = 2.2 F_N$).

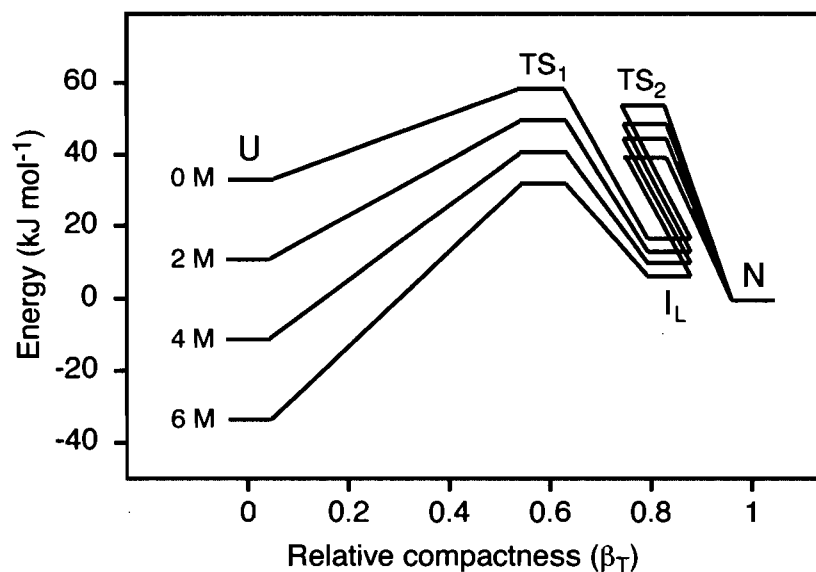


Figure 3-8. Free energy diagram for Ub folding/unfolding. The parameters obtained from Ub^{MIW} were used to calculate the free energies of the U-, I-, and the first and second transition-states (TS₁ and TS₂) relative to the N-state at various GdnHCl concentrations (see Table 3-3). A pre-exponential term of $2.4 \times 10^7 \text{ s}^{-1}$ was used to calculate the energy of the transition-states in this diagram. The x-axis reports the relative compactness of each species. The second transition-state (TS₂) of Ub is less compact than the late I-state (I_L) suggesting that some residues in the late I-state (I_L) have to get re-expose to the solvent so that the polypeptide chain can reach the last transition state. Putative early folding I-states are not illustrated in this model.

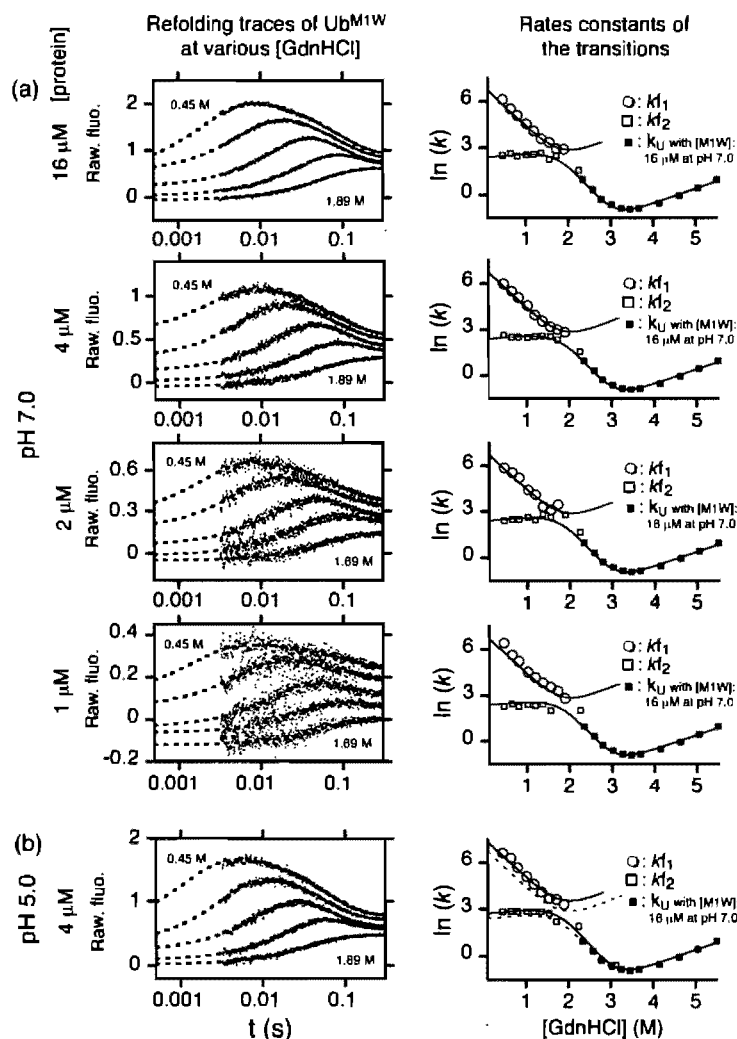


Figure 3-9. Refolding traces (left) and rate constants of the two transitions detected between 3.2 and 300 ms (right) when refolding Ub^{M1W} in different GdnHCl concentration, at different protein concentrations and at pH 7.0 (a) and 5.0 (b). Unfolding rate constants (closed squares) and the three-state on-pathway fitting of the rate constants are the same in all panels (except fitting at pH 5.0) and were taken from Figure 5. Rate constants are found to be independent of the protein concentration (down to 1 μM) and are only slightly enhanced by lowering the pH at 5.0. Refolding traces represent the average of three (for 16 μM , 4 μM or 2 μM of Ub) or six acquisitions (for 1 μM of Ub) in order to increase the signal/noise ratio.

3.9. References

1. Plaxco, K. W., Simons, K. T. & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J Mol Biol* **277**, 985-94.
2. Jackson, S. E. (1998). How do small single-domain proteins fold? *Fold Des* **3**, R81-91.
3. Fersht, A. R. (1999). *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*, Freeman, W.H., New-York.
4. Zarrine-Afsar, A., Larson, S. M. & Davidson, A. R. (2005). The family feud: do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* **15**, 42-9.
5. Paci, E., Lindorff-Larsen, K., Dobson, C. M., Karplus, M. & Vendruscolo, M. (2005). Transition state contact orders correlate with protein folding rates. *J Mol Biol* **352**, 495-500.
6. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* **324**, 359-71.
7. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953-63.
8. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat Struct Biol* **8**, 452-8.
9. Silow, M. & Oliveberg, M. (1997). Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc Natl Acad Sci U S A* **94**, 6084-6.

10. Wildegger, G. & Kiefhaber, T. (1997). Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* **270**, 294-304.
11. Gianni, S., Travaglini-Allocatelli, C., Cutruzzola, F., Brunori, M., Shastry, M. C. & Roder, H. (2003). Parallel pathways in cytochrome c(551) folding. *J Mol Biol* **330**, 1145-52.
12. Krishna, M. M., Lin, Y. & Englander, S. W. (2004). Protein misfolding: optional barriers, misfolded intermediates, and pathway heterogeneity. *J Mol Biol* **343**, 1095-109.
13. Kim, P. S. & Baldwin, R. L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* **51**, 459-89.
14. Wallace, L. A. & Matthews, C. R. (2002). Sequential vs. parallel protein-folding mechanisms: experimental tests for complex folding reactions. *Biophys Chem* **101-102**, 113-31.
15. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054-63.
16. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* **3**, 193-205.
17. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696-701.
18. Vallee-Belisle, A., Turcotte, J. F. & Michnick, S. W. (2004). raf RBD and ubiquitin proteins share similar folds, folding rates and mechanisms despite having unrelated amino acid sequences. *Biochemistry* **43**, 8447-58.
19. Went, H. M., Benitez-Cardoza, C. G. & Jackson, S. E. (2004). Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett* **567**, 333-8.

20. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
21. Crespo, M. D., Simpson, E. R. & Searle, M. S. (2006). Population of on-pathway intermediates in the folding of ubiquitin. *J Mol Biol* **360**, 1053-66.
22. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* **7**, 15-28.
23. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* **337**, 463-75.
24. Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**, 229-37.
25. Engelhard, M. & Evans, P. A. (1995). Kinetics of interaction of partially folded proteins with a hydrophobic dye: evidence that molten globule character is maximal in early folding intermediates. *Protein Sci* **4**, 1553-62.
26. Englander, S. W. & Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **16**, 521-655.
27. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. (1996). Mechanisms and uses of hydrogen exchange. *Curr Opin Struct Biol* **6**, 18-23.
28. Udgaonkar, J. B. & Baldwin, R. L. (1988). NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **335**, 694-9.
29. Roder, H., Elove, G. A. & Englander, S. W. (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* **335**, 700-4.
30. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-90.
31. Garvey, E. P., Swank, J. & Matthews, C. R. (1989). A hydrophobic cluster forms early in the folding of dihydrofolate reductase. *Proteins* **6**, 259-66.

32. Smith, C. J., Clarke, A. R., Chia, W. N., Irons, L. I., Atkinson, T. & Holbrook, J. J. (1991). Detection and characterization of intermediates in the folding of large proteins by the use of genetically inserted tryptophan probes. *Biochemistry* **30**, 1028-36.
33. Mann, C. J., Royer, C. A. & Matthews, C. R. (1993). Tryptophan replacements in the trp aporepressor from *Escherichia coli*: probing the equilibrium and kinetic folding models. *Protein Sci* **2**, 1853-61.
34. Sherman, M. A., Beechem, J. M. & Mas, M. T. (1995). Probing intradomain and interdomain conformational changes during equilibrium unfolding of phosphoglycerate kinase: fluorescence and circular dichroism study of tryptophan mutants. *Biochemistry* **34**, 13934-42.
35. Steer, B. A. & Merrill, A. R. (1995). Guanidine hydrochloride-induced denaturation of the colicin E1 channel peptide: unfolding of local segments using genetically substituted tryptophan residues. *Biochemistry* **34**, 7225-33.
36. Azuaga, A. I., Canet, D., Smeenk, G., Berends, R., Titgemeijer, F., Duurkens, R., Mateo, P. L., Scheek, R. M., Robillard, G. T., Dobson, C. M. & van Nuland, N. A. (2003). Characterization of single-tryptophan mutants of histidine-containing phosphocarrier protein: evidence for local rearrangements during folding from high concentrations of denaturant. *Biochemistry* **42**, 4883-95.
37. Jackson, S. E. (2006). Ubiquitin: a small protein folding paradigm. *Org Biomol Chem* **4**, 1845-53.
38. Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S. & Shaw, K. (1994). Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem J* **299** (Pt 1), 151-8.
39. Fersht, A. R. (2004). Phi value versus psi analysis. *Proc Natl Acad Sci U S A* **101**, 17327-8.

40. Sosnick, T. R., Dothager, R. S. & Krantz, B. A. (2004). Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc Natl Acad Sci U S A* **101**, 17377-82.
41. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-28.
42. Myers, J. K. & Oas, T. G. (2002). Mechanism of fast protein folding. *Annu Rev Biochem* **71**, 783-815.
43. Heidary, D. K., O'Neill, J. C., Jr., Roy, M. & Jennings, P. A. (2000). An essential intermediate in the folding of dihydrofolate reductase. *Proc Natl Acad Sci U S A* **97**, 5866-70.
44. Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using psi analysis. *Chem Rev* **106**, 1862-76.
45. Platt, G. W., Simpson, S. A., Layfield, R. & Searle, M. S. (2003). Stability and folding kinetics of a ubiquitin mutant with a strong propensity for nonnative beta-hairpin conformation in the unfolded state. *Biochemistry* **42**, 13762-71.
46. Bryngelson, J. D. & Wolynes, P. G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524-8.
47. Goldstein, R. A., Luthey-Schulten, Z. A. & Wolynes, P. G. (1992). Protein tertiary structure recognition using optimized Hamiltonians with local interactions. *Proc Natl Acad Sci U S A* **89**, 9029-33.
48. Shea, J. E., Onuchic, J. N. & Brooks, C. L., 3rd. (1999). Exploring the origins of topological frustration: design of a minimally frustrated model of fragment B of protein A. *Proc Natl Acad Sci U S A* **96**, 12512-7.
49. Clementi, C., Nymeyer, H. & Onuchic, J. N. (2000). Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* **298**, 937-53.

50. Briggs, M. S. & Roder, H. (1992). Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci U S A* **89**, 2017-21.
51. Park, S. H., Shastry, M. C. & Roder, H. (1999). Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat Struct Biol* **6**, 943-7.
52. Campbell-Valois, F. X. & Michnick, S. W. (2007). The transition state of the ras binding domain of Raf is structurally polarized based on Phi-values but is energetically diffuse. *J Mol Biol* **365**, 1559-77.
53. Jacob, J., Krantz, B., Dothager, R. S., Thiyagarajan, P. & Sosnick, T. R. (2004). Early collapse is not an obligate step in protein folding. *J Mol Biol* **338**, 369-82.
54. Roder, H., Maki, K., Cheng, H. & Shastry, M. C. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15-27.
55. Porschke, D. & Eigen, M. (1971). Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic--oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J Mol Biol* **62**, 361-81.
56. Tsong, T. Y. & Baldwin, R. L. (1972). A sequential model of nucleation-dependent protein folding: kinetic studies of ribonuclease A. *J Mol Biol* **63**, 453-69.
57. Tanford, C. (1970). Protein denaturation. C. Theoretical models for the mechanism of denaturation. *Adv Protein Chem* **24**, 1-95.
58. Sanchez, I. E. & Kiefhaber, T. (2003). Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J Mol Biol* **325**, 367-76.
59. Capaldi, A. P., Kleanthous, C. & Radford, S. E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* **9**, 209-16.
60. Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S., Freund, S. M., Alonso, D. O., Daggett, V. & Fersht, A. R. (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863-7.

61. Vu, N. D., Feng, H. & Bai, Y. (2004). The folding pathway of barnase: the rate-limiting transition state and a hidden intermediate under native conditions. *Biochemistry* **43**, 3346-56.
62. Chu, R., Pei, W., Takei, J. & Bai, Y. (2002). Relationship between the native-state hydrogen exchange and folding pathways of a four-helix bundle protein. *Biochemistry* **41**, 7998-8003.
63. Feng, H., Vu, N. D. & Bai, Y. (2005). Detection of a hidden folding intermediate of the third domain of PDZ. *J Mol Biol* **346**, 345-53.
64. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P., Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R. & Plaxco, K. W. (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* **14**, 602-16.
65. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* **4**, 2411-23.
66. Peterman, B. F. (1979). Measurement of the dead time of a fluorescence stopped-flow instrument. *Anal Biochem* **93**, 442-4.
67. Friel, C. T., Beddard, G. S. & Radford, S. E. (2004). Switching two-state to three-state kinetics in the helical protein Im9 via the optimisation of stabilising non-native interactions by design. *J Mol Biol* **342**, 261-73.

Chapitre 4. Article #3

Suivant la détection de l'intermédiaire tardif mal replié dans l'article #2 (Chapitre 3), et de la caractérisation de sa stabilité et de son niveau de compaction à l'aide de tryptophanes très sensibles, nous avons poussé la technique de sondage au tryptophane plus loin en élaborant un nouveau paramètre expérimental, la valeur- ω , qui rend possible la visualisation de cet intermédiaire (voir section 1.2.1.2.).

Titre : Probing the Structural Landscape of Protein Transient-States by Tryptophan Scanning Mutagenesis.

Auteurs : Alexis Vallée-Bélisle & Stephen W. Michnick

Soumission : Mai 2008

Abbreviations

Ub, mammalian Ubiquitin; Trp, tryptophan; GdnHCl, guanidine hydrochloride; T, transition; I, intermediate; N, native; U, unfolded; A, amplitude; k , rate constants; F, fluorescence;

4.1. Abstract

Proteins can exist in a variety of stable native, or transient intermediate states during folding, in response to post-translational modifications or binding to other molecules. The identification and structural characterization of transient intermediate states is a major challenge and high resolution structural methods (e.g. NMR or x-rays crystallography) do not have the temporal resolution to allow for their detection. However, fluorescence spectroscopy, using intrinsic fluorescence of tryptophan residues, can be used to detect the existence of transient intermediates for most proteins with a temporal resolution down to hundreds of nanoseconds, but with no structural resolution. Here we demonstrate that solvent exposed amino acids on a protein can be substituted for tryptophan residues with minimal perturbation of a protein stability or folding landscape. These numerous fluorescent probes can then be used to characterize the kinetics and energetics of structural transitions as well as characterizing the structures of the transient-states by comparing the fluorescence signal of a tryptophan in an unknown state to known reference states of the protein (i.e. unfolded and folded states) at many different locations. Herein, we provide the spatial characterization of a late intermediate state detected during the folding of the small protein ubiquitin by analyzing the refolding kinetic of 27 different solvent-exposed tryptophan mutants. This approach may further open the door to accurate structural determination of transient states by providing numerous structural constraints derived from the fluorescence signal of the many introduced tryptophans.

4.2. Introduction

The intrinsic fluorescence of the indole sidechain of Tryptophan (Trp) has been the classical probe to study short-lived protein transient-states because its intensity is highly sensitive to the local chemical environment¹⁻³. Generally, the characterization of protein transient-intermediate states has been conducted by using the fluorescence signal of natural or genetically engineered Trp residues, introduced at aromatic residue positions or partially buried locations, to study the effect of various mutations on the stability of this state.^{4, 5} However, in principle, one could also monitor structural changes at many defined sites along a polypeptide chain by using the fluorescence signal of Trp substituted for various solvent-exposed residues throughout the native structure of a protein (Fig. 1). In fact, Trp at solvent-exposed locations should have little effect on both the native-state stability and folding landscape while remaining highly sensitive to the detailed local chemical environment.⁶ Figure 4-1 illustrates this strategy where transient-state B (Figure 4-1A) is characterised by determining whether different Trp probes, located at various sites in the polypeptide chain (for example, sites x or y), have A-like or C-like fluorescence. For example, the fluorescence of Trp x should be mostly C-like in transient-state B because the local environment around this position is already in a C-like conformation (largest fluorescence change detected during T_1). In contrast, the fluorescence of Trp y should remain mostly A-like in transient-state B because C-like conformation is only reached during T_2 around this location (largest fluorescence change detected during T_2). Here, we used this simple strategy to probe the transient intermediate states that accumulate during the folding of the small ubiquitin protein (Ub) (76 amino acids) (Fig. 2a).^{6, 7} We recently showed that more than 99% of the Ub molecules^{ff} get trapped into a late misfolded intermediate, I_L ($\Delta G_{U-L} \sim 16$ kJ mol⁻¹), on a 10-100 milliseconds time scale (Figure 4-2b).⁶ Furthermore, there is still a debate concerning whether an early intermediate (I_E) also

^{ff} *Cis*-trapped molecules not included.

accumulates during the first milliseconds of refolding, i.e. within the dead time of a stopped-flow mixing instrument.^{6, 8-10} We therefore monitored Ub refolding using 27 different solvent-exposed Trp residues^{gg} in order to determine the structural landscape of I_L and to determine the extent of structure in the putative I_E .

4.3. Results

Surface Trps are structurally sensitive non-perturbing probes

Fluorescence intensities of a Trp in all the different states detected during Ub refolding were obtained by performing up to 25 different refolding experiments at various concentrations of denaturant (from 0.45 M to 5.86 M GdnHCl). From the end point of the refolding traces, we generated equilibrium folding/unfolding curves, which allowed us to determine both the stability of the native-state (Figure 4-3A), and the fluorescence intensity of the native and unfolded states (Figure 4-3C *top*). From the analysis of the kinetic traces, we obtained the rate constants for the formation (T_1) and rearrangement of the I_L state (T_2) (Figure 4-3B) and the fluorescence intensities of the Trp in both the I_L state (fluorescence after T_1^{hh} -Figure 4-3C *middle*) and the state preceding T_1 (F_0 : Figure 4-3C *bottom*). All of the 27 Trp detected the folding/unfolding transition (Supplementary Figure 4-1) and the unfolding free energies (ΔG^{H_2O}) obtained from the two-state fit of the equilibrium curves

^{gg} 75% of the locations most exposed to solvent in Ub.

^{hh} In the present study, the fluorescence intensity in the I_L -state, F_{IL} , should correspond to the expected fluorescence intensity reached after T_1 unless a Trp insertion dramatically reduces the fraction of I -state accumulating. For example, even a 7 kJ mol⁻¹ destabilization ($\Delta G_{U-I} \sim 5,9$ kJ mol⁻¹) would still allow $\sim 90\%$ of molecules to accumulate in the I_L .

were found to be evenly distributed around the value obtained for *wild type* Ub (28.0 kJ mol⁻¹)¹¹ (Figure 4-3a *bottom*). Twenty-four out of 27 Trps were also sensitive to both T₁ and T₂ with the exceptions of four Trps: Trp 66 only detected T₂, D32W and D39W only detected T₁, and Trp 71 could not distinguish T₁ from T₂ because their rate constants are too similar (within 3 fold) (Supplementary Figure 4-1). The rate constants for the formation and rearrangement of I_L, T₁ (open bars) and T₂ (closed bars), were not significantly affected by any of the Trp insertions (Figure 4-3b *bottom*) except for mutation E16W, which caused an enhancement of the rate of formation of the I_L-state (T₁) by a factor of 18.ⁱⁱ Finally, global fitting analysis of all equilibrium and kinetic parameters using a three-state folding model showed that the stability of I_L is relatively unaffected by Trp insertion, on average 12.9 +/- 2 kJ mol⁻¹ (see Supplementary Figure 4-2 and Table #1 for all parameters of mutants).

Trp fluorescence in transient-states

We found that the fluorescence intensity in I_L was similar to that in the native-state at many residue positions. This suggests that most of the chemical environments probed in I_L display a native-like conformation in the I-state (Figure 4-3c *middle*). For example, local environments around residues 16, 35, 45 and 60 increased fluorescence quenching of the Trp in both the I_L- and N-states compared to the denatured-state. Also, local environments around residues 48 and 53 induced the strongest Trp fluorescence intensity in both the I_L- and N-states. However, notable exceptions included Trp located at positions 57 and 59 for which the fluorescence intensity in I_L was mostly denatured-like (Trp 57: no native-like quenching; Trp 59: no native-like enhanced fluorescence intensity). Also, Trp inserted at

ⁱⁱ Trp insertion in hairpins has been reported to induce significant structural stabilization¹² and the formation of the N-terminal hairpin in Ub (β_1 and β_2) has been shown to represent a rate-limiting step in Ub folding.¹³

positions 1, 12, 42 and 73 all display higher fluorescence intensity in I_L compared to both the U- and N-states.

We also found that the fluorescence intensities of all Trp probes preceding T_1 were slightly higher than the value predicted for the denatured state (Figure 4-3c *bottom*). This small upward deviation rapidly disappeared as the denaturant concentration was increased (see Supplementary Figure 4-2). Interestingly, no Trp fluorescence quenching was observed at position 16 (first hairpin), 35 (helix C-cap), 45 (2d hairpin), 57 (3_{10} helix), and 60 (loop) suggesting that no stable native-like conformation may be formed around these positions prior the first detected refolding transition (T_1). The upward fluorescence intensity observed at low denaturant concentration may represent either a small population of non-native-like intermediate that accumulates during the mixing time of the stopped-flow instrument or a change in the unfolded state conformation due to lower concentration of denaturant.¹⁵

Fluorescence-based structural description

While it is not possible to deduce precise structural constraints based on local chemical effects on the fluorescence quantum yield of Trp,¹⁶ it is reasonable to argue that variation in fluorescence intensity stems from the proximity of quenching groups in the vicinity of the Trp.¹⁷ The difference in fluorescence intensity between a transient-state i , and a reference state j for a given Trp residue at position k , $\Delta F_{j-i}(k)$, should therefore provide a good indicator of local structural variation between both states in proximity to residue k . Furthermore, a better description of a transient-state conformation could in principle be obtained by comparing Trp fluorescence intensity in this state to the Trp fluorescence intensities obtained for two known states located prior and after the transient-state in a specific reaction scheme (Figure 4-1). In the case of a folding transient-state, both the native and the unfolded-states could be used as reference conformations. The ratio, ΔF_L ,

$I_U(k)/\Delta F_{N-U}(k)$, which we call the omega (ω)-value allows for the quantification of these trends: a ratio smaller than 0.5 would indicate that for a given probe at position k in the protein, the intermediate-state (I) fluorescence is closer to that of the unfolded-state (U) than to that of the native-state (N). These values can then be mapped onto the native structure in order to visualize which parts of the structure are already formed in the intermediate, or remain in the denatured or other non-native states (Figure 4-4).

Figure 4-4 presents two distinct fluorescence-based structural descriptions of I_L detected during Ub refolding using either $\Delta F_{N-I}(k)$ or the ω -value. First, we observe that three of the five Trps that displayed the largest ΔF_{N-I} , Trp57, Trp59 and Trp60, are located in or nearby the last 3/10 helix (Figure 4-4 *left*). These positions also displayed ω -values close to zero which suggests that local environment around the 3/10 helix is mainly denatured-like in I_L (Figure 4-4 *right*). The two other Trps are located at positions 71 and 12; their fluorescence being denatured-like and non-native-like in I_L , respectively (ω -values: 0.02 and 2.2). Trp 66 also displayed denatured-like fluorescence values in the I_L ($\omega = 0.41$). Finally, with Trp 12, three others Trp provided much higher fluorescence intensity in I_L than in both the U- and N-states ($\omega \gg 1$) providing evidence for the presence of structurally distinct chemical environments around these positions: residue 73 at the extremity of the last strand (β_5), and residues 1 (β_1) and 42 (β_3) both in interacting with strand β_5 . In summary, seven of the nine positions that displayed the highest ΔF_{N-I} are located between residues 57 and 73; the two others (1 and 12) are in close structural contact with this segment in the native-state. Also, eight of the nine positions that displayed apparent denatured-like Trp fluorescence intensity in the I_L -state were also located in this segment. These results suggest that the last C-terminal segment of Ub, which contains the 3/10 helix and β_5 , may not be properly packed in I_L . Additional data from the literature supports this model. For example, a NMR H/D pulse-exchange study performed on *wild type* Ub refolding showed that amide H-bonds of residues 59, 61, and 69 are protected at a lower rate than the rest of the amide protons upon refolding.¹⁸ Also, NMR structural

evidences obtained on additional stable forms of Ub induced by high pressure,^{19, 20} low temperature,²⁰ or various solvent conditions²¹ suggest that the structure of Ub possesses a considerable degree of structural plasticity in order to compensate for the opening of strand β_5 . This feature, which was proposed to play an important role in Ub function [the C-terminus end of Ub is known to interact with three enzymes]⁷, also provides a suitable explanation for the remarkably high stability of I_L.

4.4. Discussion

The global Trp scanning mutagenesis strategy described here provides a minimally perturbing method for extracting kinetic and structural information on protein transient states at time-scales down to hundreds of nanoseconds. It is a simple, accessible and inexpensive approach to characterize any structural or allosteric transitions that occur as a result of a chemical or physical perturbation, or post-translational modifications.^{jj} It could also be used to evaluate mechanisms underlying structural and dynamic coupling of structurally distant residues that show high evolutionary conservation among homologues²² or to screen small molecule (i.e. non-competitive inhibitors) that induce unique conformational changes at distant sites.^{23, kk} Furthermore, a variation of this approach using fluorescent analogues of tryptophan²⁴ or other fluorophores could also be applied to monitor structural changes *in vivo*²⁵ and serve to design novel biosensors for the detection of any structural or chemical modification of a protein.^{26, 27}

^{jj} It may be easily applied to many proteins since Trp is among the most infrequent amino acid and many proteins contain only one or two residues.

^{kk} For example, by choosing a molecule with the desired effect as a lead, Trp scanning could serve to map out the conformation change on the protein surface at low resolution, while a lead optimization strategy could follow in which chemical modifications could be selected for providing the lowest transition barrier to induction of the conformation change.

The Trp scanning mutagenesis strategy can also complement or aid in other approaches to study protein transient states. For example, in order to be efficient, perturbation-based strategies like Φ -value analysis⁵ require probes that are sensitive to all protein transition. Also, the fact that fluorescence spectroscopy is effective over the nanomolar-to-millimolar concentration range and nanosecond-millisecond time scale could also allow experimentalists to bridge the information gap between what needs to be learned about structures that occur rapidly at *in vivo* concentrations and what is learned from higher resolution structural techniques, such as nuclear magnetic resonance (NMR) spectroscopy or X-ray diffraction, which require high concentrations of protein and much longer acquisition times.²⁴ Alternatively, Trp scanning strategy could also be used to predict residues that should be mutated to selectively destabilize a native-state in order to populate a relevant kinetic intermediate-states at equilibrium.^{28, 29, 11} Finally, Trp scanning mutagenesis could be used to catalogue effects of different chemical environments on the fluorescence of Trp allowing for a better understanding of how different chemical groups in different contexts effect fluorescence.^{16, mm} This would provide more precise structural constraints that could be used for modeling transient states in computer simulations.³¹ Such studies could result in a knowledge of unprecedented detail concerning the ensembles of structures that can exist in transient states

¹¹ For example, in the case of Ub, one should try mutating residues included between positions 57-66 since this segment displays a very distinct conformation in the I_L and N-states.

^{mm} If one wishes to get more information out of the fluorescence signals one could use the picosecond/nanosecond time-resolved total intensity and anisotropy decay which is the most important fluorescence parameter in terms of providing detailed physical information.³⁰

Acknowledgements

The authors acknowledge Tobin Sosnick, and F.X. Campbell-Valois and Peter Richardson for helpful discussions, M. Fyfe for sequencing, and Jeffrey W. Keillor for providing access to the stopped-flow apparatus. This work was supported by the NSERC of Canada (Grant 194582-00). SWM holds the Canada Research Chair in Integrative Genomics.

4.5. Table

Table 4-1. Thermodynamic and kinetic parameters obtained for all Trp mutants. See Supplementary Figure 4-1 (ΔG_{UN} , m_e , k_1 , k_2 , k^{2M} , m_F^{2M}) and Supplementary Figure 4-2 (ΔG_{UI}) for complete details.

Mutation	Location	$\Delta G_{UN}^{H_2O a}$	m_e^b	$k_1^{0.45M c}$	$k_2^{0.45M c}$	$k^{2M d}$	$m_F^{2M e}$	$\Delta G_{UI}^{H_2O a, f}$
M1W	β_1	36.8 (5.8)	11.0 (1.8)	316 (6)	12.1 (0.1)	5.7 (0.3)	-7.9 (0.6)	16.8
F4W	β_1	27.8 (1.6)	10.1 (0.6)	121 (5)	20.3 (0.5)	2.4 (0.1)	-6.1 (0.2)	10.5
L8W	Turn I (i+1)	23.7 (2.2)	6.7 (0.6)	324 (11)	7.1 (0.6)	3.7 (0.1)	-6.8 (0.1)	14.8
T12W	β_2	29.6 ^k	-	90 (11) ^f	41 (3)	10.7 (1.2)	-7.7 (1.6)	13.2
E16W	β_2	34.1 ^k	-	3000 ^g	41 (7)	30.9 (1.8)	-6.7 (0.2)	16.4
S20W	Turn I (i+2)	23.0 (3.1)	7.2 (1.0)	165 (14)	26 (2)	6.6 (0.4)	-7.2 (0.5)	-
E24W	α -Helix	33.5 ^k	-	205 (16)	53 (10)	2.6 (0.1)	-6.7 (0.2)	13.3
A28W	α -Helix	35.2 (4.6)	10.8 (1.4)	162 (6)	17.6 (1.8)	2.5 (0.1)	-6.7 (0.1)	13.5
D32W	α -Helix	26.0 ^k	-	188 (8)	-	2.8 (0.1)	-7.4 (0.3)	-
G35W	Helix C-cap	19.5 (3.6)	-	120 (5)	30.1 (1.3)	0.6 (0.2)	-7.7 (0.9)	5.3
D39W	3_{10} Helix	35.4 ^k	-	157 (23)	-	3.9 (0.1)	-6.5 (0.3)	-
R42W	β_3	29.2 ^k	-	142 ^f (6)	20.4 (0.5)	5.4 (0.4)	-7.1 (0.6)	15.8
F45W	Turn IV (i+1)	28.7 (2.8)	9.5 (0.9)	178 (3)	15.6 (0.8)	3.3 (0.1)	-6.5 (0.2)	12.5
K48W	β_4	24.1 (1.9)	8.3 (0.6)	295 (12)	18.9 (0.6)	4.7 (0.2)	-7.6 (0.2)	14.7
E51W	Turn I (i)	27.0 (1.7)	9.6 (0.6)	136 (4)	12.6 (1.9)	2.9 (0.1)	-7.6 (0.3)	12.9
G53W	Turn I (i+2)	30.3 (1.9)	9.0 (0.6)	323 (9)	17.9 (0.4)	4.7 (0.2)	-6.6 (0.2)	13.5
R54W	Turn I (i+3)	32.4 (7.5)	10.2 (2.4)	256 ^g	8.9 (1.0)	4.3 (0.2)	-6.1 (0.2)	15.3
S57W	3_{10} Helix	27.9 (1.7)	9.2 (0.5)	95 (14)	13.6 (0.2)	3.4 (0.1)	-5.8 (0.2)	10.1
Y59W	3_{10} Helix	25.6 (6.2)	12.1 (2.8)	310 ^g	17.4 (0.2)	1.9 (0.6)	-6.0 (1.4)	13.0
N60W	Turn	25.0 ^k	-	390 ^g	7.8 (1)	2.4 (1.4)	-7.8 (5.1)	13.0
Q62W	Turn II (i)	29.6 ^k	-	105 (11)	7.1 (0.5)	4.2 (0.3)	-4.8 (0.3)	12.6
E64W	Turn II (i+2)	22.3 (3.6)	6.2 (1.1)	136 (11)	14.4 (0.3)	6.2 (0.3)	-5.6 (0.2)	14.9
T66W	β_5	27.0 (6.1)	9.4 (2.1)	-	13.9 (0.2)	3.2 (0.1)	-5.1 (0.2)	10.2
H68W	β_5	24.6 (4.5)	6.7 (1.3)	92 (12)	15.8 (1.1)	6.2 (0.4)	-5.8 (0.3)	13.8
V70W	β_5	28.1 (2.8)	9.4 (0.9)	140 (6)	11.6 (0.8)	2.3 (0.1)	-7.2 (0.4)	11.4
L71W	β_5	26.9 (5.3)	8.6 (1.7)	123 (1)	-	1.9 (0.1)	-7.0 (0.2)	7.3 ^j
L73W	Loose	28.2 (8.3)	8.5 (2.5)	282 (16)	42.8 (0.8)	2.7 (0.1)	-6.7 (0.3)	8.8
T66W/F45W	-	24.1 (8.4)	9.5 (3.2)	303 (10)	9.5 (0.5)	4.2 (0.6)	-7.9 (1.6)	19.5
L71W/F45W	-	30.9 ^k	-	163 (28)	83 (7)	1.7 (0.1)	-6.8 (0.4)	7.3
Average	-	28.2 (3.2)	9.0 (1.2)	202 ^h (77)	24 (13)	4.8 (2.6)	-6.8 (0.6)	12.9 (2.2)

See symbol legend next page.

^a The free energies of folding (ΔG) are in kJ mol^{-1} .

^b The m_e -values are in $\text{kJ mol}^{-1} \text{M}^{-1}$.

^c Rate constant (s^{-1}) of the two folding transitions detected between 3.2 and 300 ms of refolding time in 0.45 M GdnHCl.

^d Rate constant (s^{-1}) of the unique folding transition detected at 2.0 M GdnHCl.

^e The kinetic m_F -value for folding rate above 2M GdnHCl were multiplied by the factor RT (gas constant times temperature).

^f Strong roll over for k_1 at low concentration of denaturant (Ub^{T12W} and Ub^{R42W}).

^g k_1 too fast at 0.45 M. Value extrapolated using m_{UI} determined for Ub^{M1W} .⁶

^h Average k_1 was determined without Ub^{E16W} .

ⁱ Average m_i and m_{IN} obtained from the mutants in bold were used to estimate the intermediate stability for the other mutants.

^j Value taken from the double Trp mutant analysis (L71W/F45W).

^k $\Delta G_{\text{UN}}^{\text{H}_2\text{O}}$ obtained by using the average m_e value.

4.6. Figures

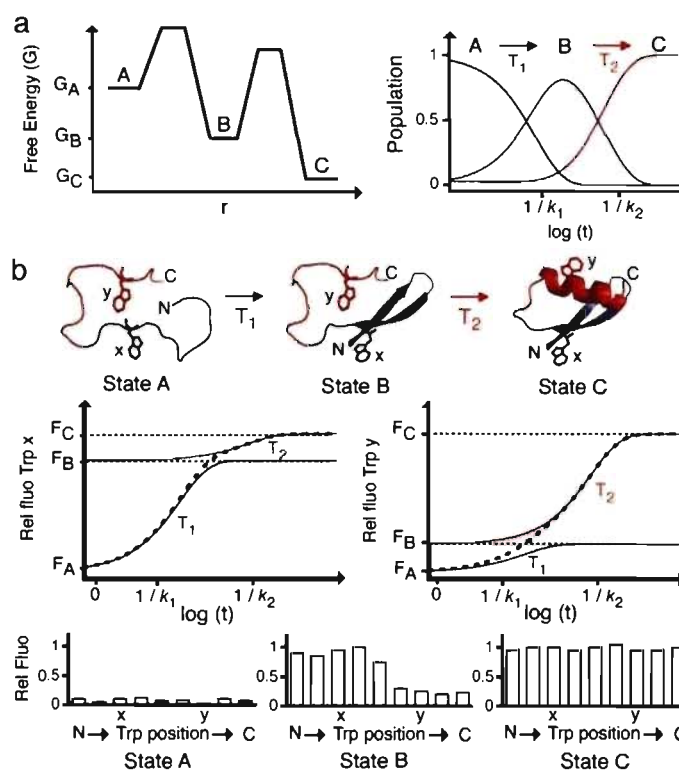


Figure 4-1. Characterizing protein transient-states using Trp scanning mutagenesis. a) Transient-state B will be highly populated during a kinetic experiment (left) if its rate of rearrangement into state C is slower than its rate of formation ($k_2 < k_1$) and if it is more stable than state A ($G_B < G_A$ —right). b) Under these conditions, the fluorescence intensity of a tryptophan (Trp) can be determined at various solvent-exposed positions in the polypeptide (x, y, ...; two Trps are shown for explanation). In a real experiment, a series of individual Trp mutants are created) in protein states A, B, and C (F_A , F_B , and F_C) via a simple modeling of the kinetic trace. Trp fluorescence in state B should display an intensity, F_B , that is similar to F_C if the local environment around Trp has already reached its final conformation (*left*: residue position x is already in a hairpin in state B), or will remain similar to F_A if the local environment around Trp hasn't significantly changed following the first transition T_1 (*right*: residue position y is not yet in a α -helix conformation in state B).

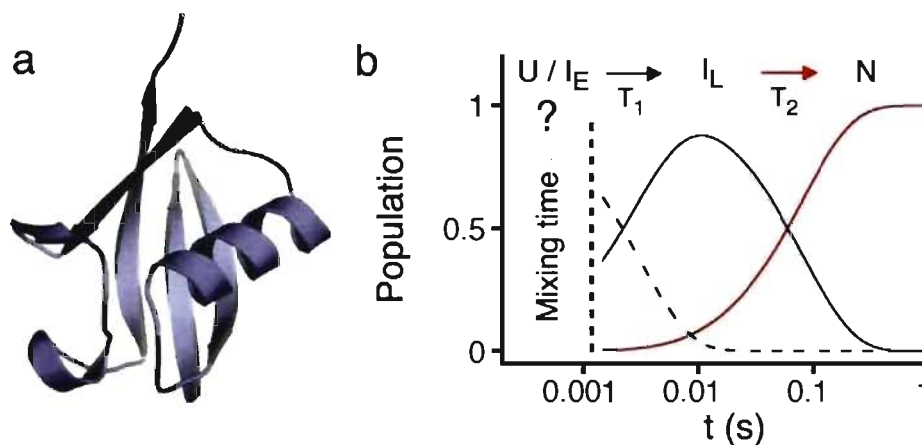


Figure 4-2. Transient intermediate states detected during ubiquitin (Ub) refolding. a) Ribbon representation of Ub structure ($\beta\beta\alpha\beta\beta\alpha\beta$ topology) (pdb code: 1UBQ)³². b) Kinetics of appearance/disappearance of a late misfolded state (I_L) (blue line) detected during the refolding of Ub^{M1W} in 0.45M GdnHCl.⁶ Up to 99.9% of Ub molecules are expected to accumulate in the I_L -state ($\Delta G_{U-I_L} = -16.8$ kJ mol⁻¹) allowing therefore the direct characterization of Trp fluorescence in this state. There is still a debate concerning whether I_L forms directly from the unfolded-state (U) or *via* an earlier folding intermediate (I_E) formed during the dead time of the mixing instrument.^{8, 9, 33}

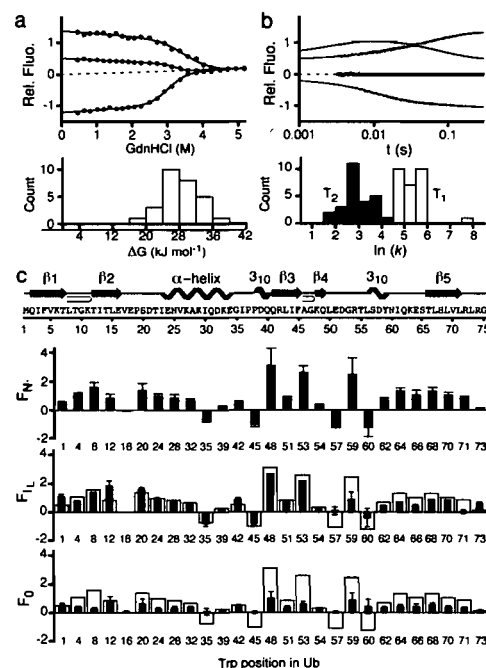


Figure 4-3. Equilibrium and kinetic characterization of the different Ub Trp mutants. a) *top*: equilibrium unfolding curves for three typical Trp mutants (blue: M1W, red: F45W, green: E64W) obtained by taking the fluorescence endpoint (total fluorescence intensity above 320 nm -cut-off filter) of various folding reactions performed at different concentrations of GdnHCl (50 mM NaHPO₄, pH 7.0 at 30°C). Protein concentration was set to 15 μ M. The unfolding curves were normalized for similar denatured-state fluorescence intensity and fitted well to a two-state unfolding model;³⁴ *bottom*: distribution of the unfolding free energy of all 27 single Trp mutants obtained from the fitting of the equilibrium curves (average free energy = 28.3 ± 3.2 kJ mol⁻¹); b) *top*: bi-exponential kinetic refolding traces of the three typical single Trp mutants obtained following a rapid GdnHCl jump from 5.0 M to 0.45 M and by monitoring total fluorescence intensity above 320 nm (cut-off filter). *Wild type* Ub with no Trp displays no fluorescence changes (black dots). Four mutants displayed only one kinetic transition because either their I- and N-states (Ub^{D32W} and Ub^{D39W}), or their U and I-states (Ub^{T66W} and Ub^{L71W}) display similar fluorescence intensities (see Supplementary Figure 4-1). *Bottom*: distribution of k_1 (black bars) and k_2 (white bars) obtained from the kinetic curves at 0.45 M GdnHCl. c) Relative fluorescence intensity of Trp at 27 different solvent-exposed locations in Ub in the native-state (*top*), in the late I-state (F_I ; *middle*), and at the initiation of the folding reaction (F_0 ; *bottom*) in 0.45 M GdnHCl (N-state fluorescence also provided for reference with white rectangles). The expected fluorescence intensity of the unfolded state at 0.45 M GdnHCl was taken as a reference and set to zero for all mutants. See Supplementary Figure 4-1 and 4-2 for complete details concerning the folding/unfolding analysis of all Trp mutants.

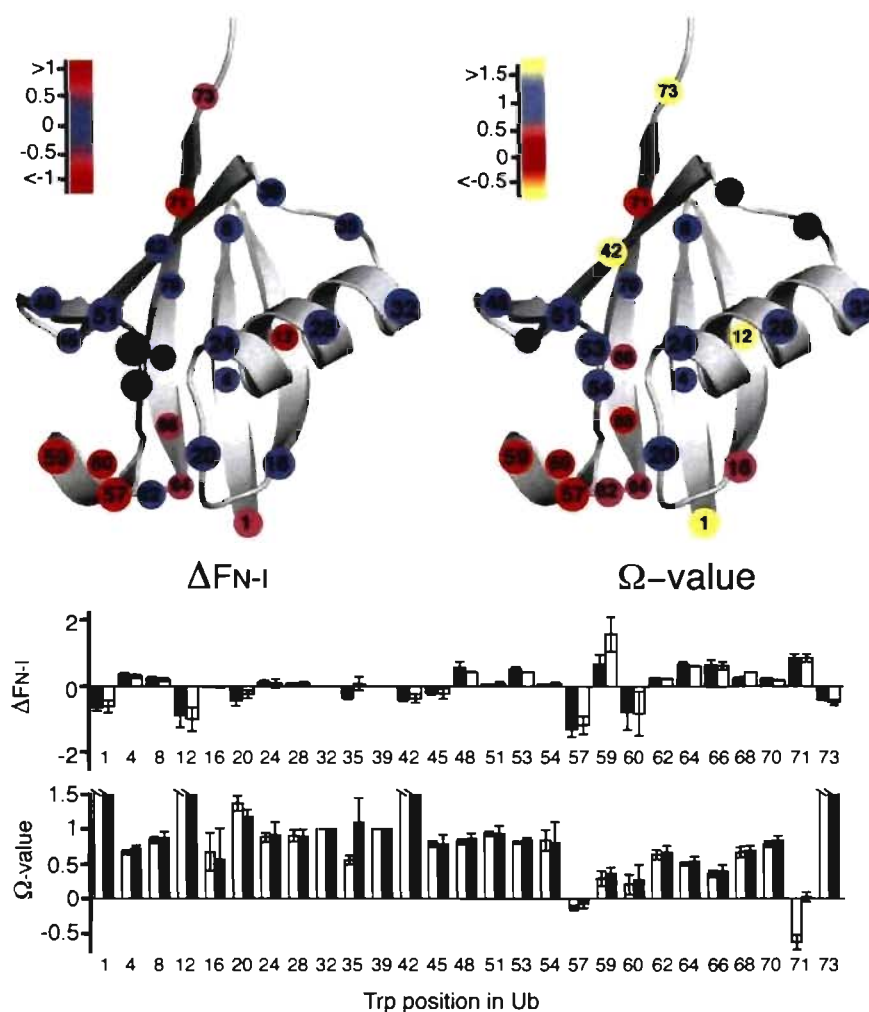
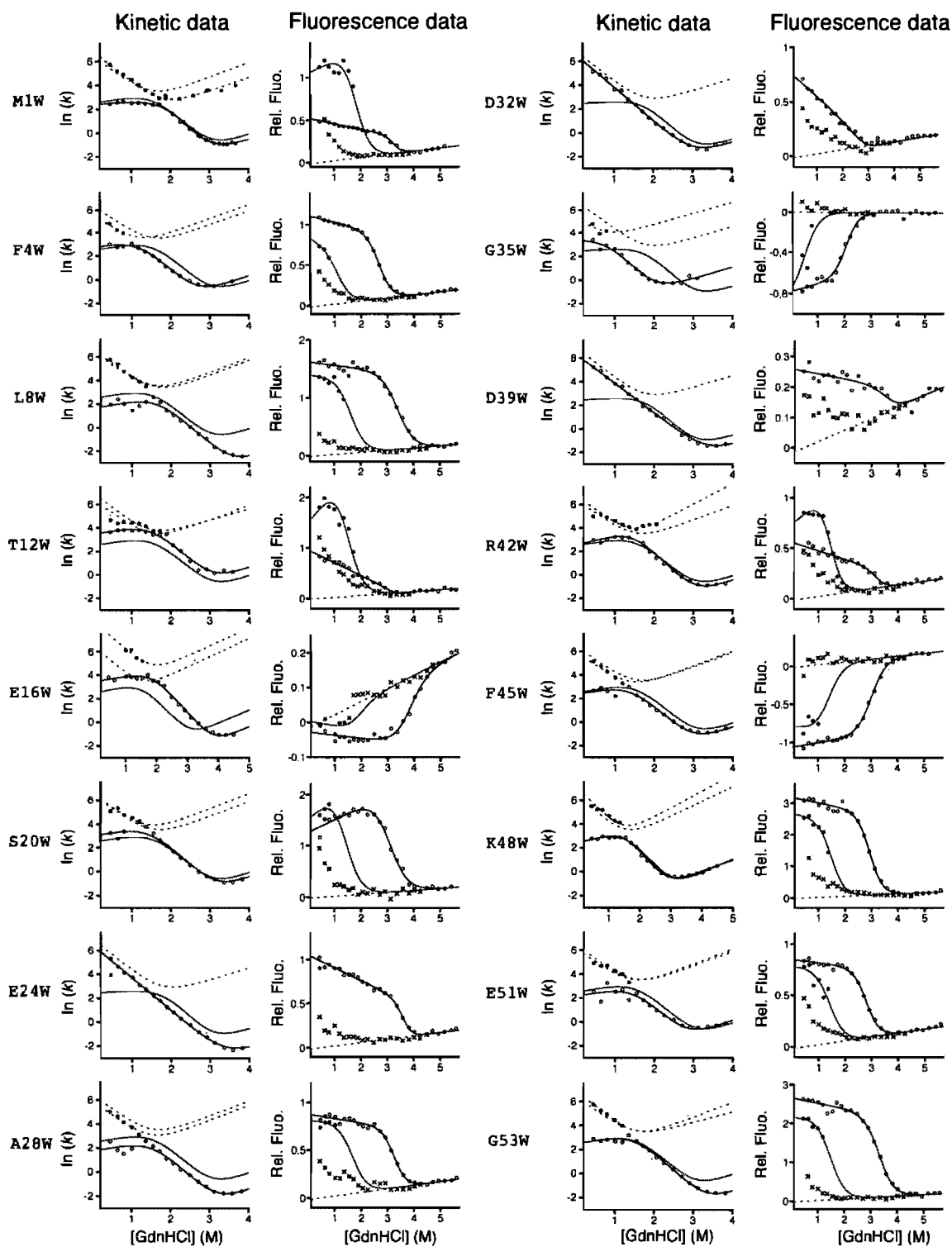


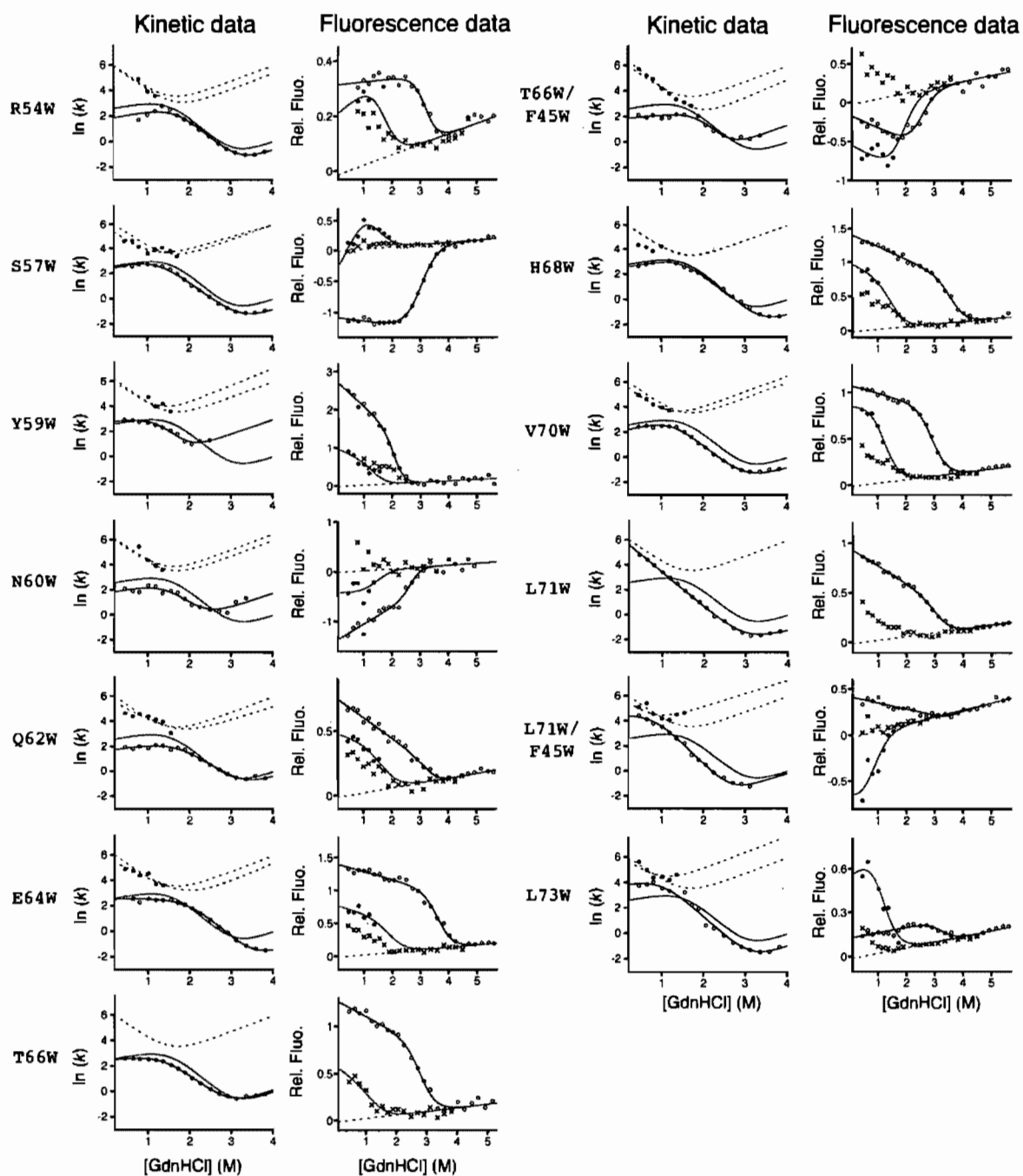
Figure 4-4. Fluorescence-based structural representation of the Ub late folding intermediate. *Left:* difference in fluorescence intensity between the native (N) and the late folding intermediate (I_L) at residue location k ($\Delta F_{N-I}(k)$). *Right:* ω -value determined at different residue positions, $\Delta F_{I-U}(k)/\Delta F_{N-U}(k)$, where $\Delta F_{i-j}(k)$ is the difference in fluorescence between state i and j for residue k ; a ratio smaller than 0.5 indicates that for a given probe at position k in the protein, the intermediate-state fluorescence (I) is closer to that of the unfolded-state (U) than to that of the native-state (N). *Bottom:* ΔF_{N-I} and ω -value determined using the refolding trace in 0.45 M GdnHCl (white rectangle) or by using the simultaneous global fitting analysis of all equilibrium and kinetic parameters at various concentration of GdnHCl (see Supplementary Figure 4-2).

Supplementary Figure 4-1. Equilibrium curves, folding traces at 0.45 M GdnHCl, and rate of the apparent “two-state” folding transition detected above 2 M GdnHCl for all Trp mutants. *Left:* Equilibrium curves were generated by plotting the fluorescence end points of all kinetic traces *versus* GdnHCl. All curves were well fitted using a “two state” folding/unfolding model (see $\Delta G_{\text{UN}}^{\text{H}_2\text{O}}$ and m_e in Table #1).³⁴ All fluorescence data were normalized for similar denatured-state fluorescence intensity between 2.5 M and 5.5 M GdnHCl (set to 0 in absence of denaturant). *Middle:* 24 out of 27 Trp detected both T_1 and T_2 (double exponential fitting –see $k_1^{0.45\text{M}}$ and $k_2^{0.45\text{M}}$ in Table #1). The refolding traces of the four other Trp mutants could be well fitted with a single exponential function: Trp 66 only detected T_2 , D32W and D39W only detect T_1 , and Trp 71 could not distinguish T_1 from T_2 because their rate constants are too similar (within 3 fold –see Table #1). A double Trp mutant strategy⁶ was used in order to detect the other missing transition (see mutants T66W-F45W and L71W-F45W). *Right:* Observed folding rate constant of the unique transition detected above 2 M GdnHCl *versus* GdnHCl (see $k^{2\text{M}}$ and $m_F^{2\text{M}}$ in Table #1). All curves were fitted using a “two-state” folding model (chevron curve).³⁴ Dotted line corresponds to the average folding kinetic behaviour.

Supplementary Figure 4-2. Global fitting analysis of the kinetic and thermodynamic parameters obtained for all Trp mutants. *Kinetic data:* the observed rate constant for formation of I_L (k_1 ; filled circles) was fit to a chevron curve (black dotted line) thus providing information on k_{UI} , k_{IU} , m_{UI} , and m_{IU} (and m_I and ΔG_{UI} , indirectly). Data from double jump experiments⁶ were also provided for mutant M1W in order to show the complete folding/unfolding behavior of I_L . The observed rate constant for the rearrangement of I_L (k_2 ; filled circles) was fit to a three-state on-pathway model (black line) in order to obtain k_{IN} , k_{NI} , m_I and ΔG_{UI} .⁶ Red lines correspond to the average folding kinetic behavior. *Fluorescence data:* the fluorescence end points of all kinetic traces (open circles) were fitted *versus* GdnHCl using a “two state” folding/unfolding model (black line) in order to obtain an estimate of m_N and ΔG_{UN} .³⁴ The fluorescence values after T_1 (F_1 ; filled circles) were fitted *versus* GdnHCl using a “two state” folding/unfolding model (red line) in order to obtain an estimate of m_I and ΔG_{UI} . The stability of I_L was therefore determined from the best simultaneous fit of k_1 , k_2 and F_1 over all concentrations of denaturant (see $\Delta G_{UI}^{H_2O}$ in Table #1).⁶ Fluorescence of the protein state prior to T_1 is shown with x. Standard deviations were obtained from the best fit of the data.

Additional information on fitting procedure. 10 mutants provided sufficient kinetic resolution of k_2 *versus* concentration of GdnHCl in order to allow for a precise estimation of m_{IN} (mostly Trp inserted in C-terminal portion: W1, W42, W48, W53, W57, W62, W64, W66, W70, W66/W45). The average m_{IN} ($1.5 \text{ kJ mol}^{-1} \text{ M}^{-1}$) and m_I (8.9 kJ mol^{-1}) obtained from these mutants were used in order to estimate ΔG_{UI} for the other mutants. In all simulations, the unfolding rate denaturant dependency of the native-state, m_{NI} , was set to $2.85 \text{ kJ mol}^{-1} \text{ M}^{-1}$ (average value obtained for five mutants in previous study).⁶ Also, the unfolding rate denaturant dependency of the intermediate state, m_{IU} , was set to the value obtained for Ub^{M1W} ($3.1 \text{ kJ mol}^{-1} \text{ M}^{-1}$)⁶ when the folding rate denaturant dependency of the intermediate state m_{UI} could not be estimated with enough precision. For more detail on the global fitting procedure used please refer to reference 6.





4.7. References

1. Teale, F. W. & Weber, G. (1957). Ultraviolet fluorescence of the aromatic amino acids. *Biochem J* **65**, 476-82.
2. Weber, G. (1960). Fluorescence-polarization spectrum and electronic-energy transfer in proteins. *Biochem J* **75**, 345-52.
3. Royer, C. A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chem Rev* **106**, 1769-84.
4. Matouschek, A., Kellis, J. T., Jr., Serrano, L., Bycroft, M. & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440-5.
5. Fersht, A. R. (1999). *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*, Freeman, W.H., New-York.
6. Vallee-Belisle, A. & Michnick, S. W. (2007). Multiple tryptophan probes reveal that ubiquitin folds via a late misfolded intermediate. *J Mol Biol* **374**, 791-805.
7. Jackson, S. E. (2006). Ubiquitin: a small protein folding paradigm. *Org Biomol Chem* **4**, 1845-53.
8. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696-701.
9. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
10. Went, H. M., Benitez-Cardoza, C. G. & Jackson, S. E. (2004). Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett* **567**, 333-8.
11. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054-63.
12. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. (2001). Tryptophan zippers: stable, monomeric beta -hairpins. *Proc Natl Acad Sci U S A* **98**, 5578-83.

13. Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**, 229-37.
14. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* **337**, 463-75.
15. Jennings, P. A. (1998). Speeding along the protein folding highway, are we reading the signs correctly? *Nat Struct Biol* **5**, 846-8.
16. Callis, P. R. & Liu, T. (2004). Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J Phys Chem B* **108**, 4248-4259.
17. Demchenko, A. P. (1986). *Ultraviolet spectroscopy of proteins*, Springer-Verlag, New-York.
18. Briggs, M. S. & Roder, H. (1992). Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci U S A* **89**, 2017-21.
19. Kitahara, R., Yamada, H. & Akasaka, K. (2001). Two folded conformers of ubiquitin revealed by high-pressure NMR. *Biochemistry* **40**, 13556-63.
20. Kitahara, R. & Akasaka, K. (2003). Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding. *Proc Natl Acad Sci U S A* **100**, 3167-72.
21. Pan, Y. & Briggs, M. S. (1992). Hydrogen exchange in native and alcohol forms of ubiquitin. *Biochemistry* **31**, 11405-12.
22. Suel, G. M., Lockless, S. W., Wall, M. A. & Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* **10**, 59-69.
23. Hardy, J. A. & Wells, J. A. (2004). Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* **14**, 706-15.
24. Ross, J. B., Szabo, A. G. & Hogue, C. W. (1997). Enhancement of protein spectra with tryptophan analogs: fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. *Methods Enzymol* **278**, 151-90.

25. Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W. & Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **103**, 9785-9.
26. Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-24.
27. Wilson, G. S. & Gifford, R. (2005). Biosensors for real-time in vivo measurements. *Biosens Bioelectron* **20**, 2388-403.
28. Sanz, J. M. & Fersht, A. R. (1993). Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry* **32**, 13584-92.
29. Spence, G. R., Capaldi, A. P. & Radford, S. E. (2004). Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215-26.
30. Beechem, J. M. (1997). Picosecond fluorescence decay curves collected on millisecond time scale: direct measurement of hydrodynamic radii, local/global mobility, and intramolecular distances during protein-folding reactions. *Methods Enzymol* **278**, 24-49.
31. Vendruscolo, M. & Dobson, C. M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. *Philos Transact A Math Phys Eng Sci* **363**, 433-50; discussion 450-2.
32. Vijay-Kumar, S., Bugg, C. E. & Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* **194**, 531-44.
33. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* **324**, 359-71.
34. Maxwell, K. L., Wildes, D., Zarrine-Afsar, A., De Los Rios, M. A., Brown, A. G., Friel, C. T., Hedberg, L., Horng, J. C., Bona, D., Miller, E. J., Vallee-Belisle, A., Main, E. R., Bemporad, F., Qiu, L., Teilum, K., Vu, N. D., Edwards, A. M., Ruczinski, I., Poulsen, F. M., Kragelund, B. B., Michnick, S. W., Chiti, F., Bai, Y., Hagen, S. J., Serrano, L., Oliveberg, M., Raleigh, D. P., Wittung-Stafshede, P.,

Radford, S. E., Jackson, S. E., Sosnick, T. R., Marqusee, S., Davidson, A. R. & Plaxco, K. W. (2005). Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* **14**, 602-16.

Chapitre 5. Discussion

Au cours de l'introduction, nous avons présenté quelques exemples illustrant le rôle primordial qu'occupent désormais les recherches fondamentales sur le mécanisme de fonctionnement ou de repliement des protéines dans les domaines de la santé et des biotechnologies. Nous avons également vu que l'élucidation du mécanisme de fonctionnement ou de repliement d'une protéine nécessite la description structurale et énergétique complète de tous les états intermédiaires et de transitions empruntées par la protéine au cours du dit mécanisme. Diverses approches permettent actuellement d'obtenir plusieurs types d'informations structurales de haute résolution sur les états de protéines présents à l'équilibre (principalement les états natifs et dépliés) (Table 1-1). En revanche, les outils mis à la disposition du biochimiste pour caractériser les états transitoires de protéines de courte durée de vie demeurent relativement limités.

Les résultats présentés au sein du Chapitre 2 illustrent bien les limitations inhérentes à la caractérisation des états transitoires détectés durant le repliement des protéines. L'objectif premier de cet article visait à vérifier si le DLR de *raf* et l'ubiquitine, deux protéines possédant des structures similaires, sans identité de séquence apparente (Figure 2-1), se replient via des mécanismes similaires. Cette comparaison a été réalisée dans le but de tester la théorie voulant que le mécanisme de repliement d'une protéine est davantage déterminé par sa structure que par les fins détails structuraux de sa séquence acides aminés.¹ Nos résultats ont, en effet, permis de déterminer que le DLR de *raf* ainsi qu'ubiquitine se replient toutes deux via quatre transitionsⁿⁿ de vitesses similaires (Figure 2-7). La première transition détectée lors du repliement de ces deux protéines, T₁, montrent les caractéristiques d'une transition de type « deux états » (Figure 2-6). De plus, la vitesse

ⁿⁿ Quatre transitions détectées entre 3.2 ms et 10 s.

de cette transition varie de manière similaire chez les deux protéines suivant une diminution de température, l'ajout de sulfate de sodium, ou lorsque différentes mutations d'acides aminés sont réalisées à une position équivalente dans le cœur hydrophobe des deux protéines (Figure 2-11).^{oo} Des expériences de double saut rapides « dépliement-repliement » ont également permis de suggérer que les troisièmes et quatrièmes transitions détectées durant le repliement de ces protéines, T_3 et T_4 , semblent être liées à des processus d'isomérisation *cis-trans* du lien peptidique de résidus non-prolyl³ et prolines,⁴ respectivement (Figure 2-7). Toutefois, aucune information additionnelle n'a pu être obtenue concernant la nature de la deuxième transition détectée durant le repliement de chacune de ces deux protéines (T_2). Ces dernières représentent-elles des intermédiaires séquentiels dans la voie de repliement des deux protéines ou simplement des fractions de protéines se repliant plus lentement (Figure 1-8A)?

Au cours du Chapitre 3, les deux premières transitions détectées durant le repliement d'ubiquitine ont ainsi été caractérisées à l'aide de cinq résidus tryptophanes localisés à différentes positions exposées au solvant dans la structure (voir "Utilisation de plusieurs mutants tryptophanes pour détecter et caractériser les états intermédiaires transitoires à la section 1.2.1.1). Cette étude a permis de démontrer que l'amplitude de ces deux transitions est radicalement différente dépendamment de la position du tryptophane utilisé pour sonder le repliement (Figure 3-5). Deux tryptophanes, Trp1 et Trp57, ont d'ailleurs permis de démontrer de manière directe que T_1 ne représente pas une transition à deux états de type É.D.→É.N., mais bien la formation d'un intermédiaire séquentiel tardif mal replié qui participe au mécanisme de repliement de la protéine (É.D.→É.I.). Au cours de cette étude, nous avons également montré, qu'il était facile d'utiliser des combinaisons de tryptophanes pour rendre la détection d'un intermédiaire transitoire évidente. Par

^{oo} Mon collègue, le Dr. François-Xavier Campbell-Valois a d'ailleurs récemment publié une étude comparative complète des deux premières transitions de type « deux états » détectées chez ces deux protéines en utilisant la méthode d'analyse par valeur- Φ .²

exemple, une combinaison de tryptophanes aux positions 45 et 66 permet d'obtenir un mutant affichant des amplitudes opposées durant T_1 et T_2 (Figure 3-6). Finalement, l'énergie ainsi que le degré de compaction de l'intermédiaire tardif ont été caractérisés de manière directe à l'aide de doubles sauts rapides « repliement-dépliement » en utilisant la fluorescence du tryptophane inséré à la position 1, qui rend possible la détection efficace de l'intermédiaire (Figure 3-7).

Enfin, le chapitre 4 rapporte la caractérisation de la structure de l'intermédiaire tardif détecté durant le repliement d'ubiquitine via une approche par sondage global au tryptophane, telle qu'introduite à la section 1.2.1.2. (« Visualiser la structure d'un intermédiaire transitoire par sondage global au tryptophane »). À partir de notre nouveau paramètre expérimental, la valeur- ω , nous avons déterminé que la section comportant les résidus 1 à 54 d'ubiquitine (deux premières épingles et hélice- α) affiche une conformation de type native au niveau de l'intermédiaire tardif. Au contraire, la section 57-66 d'ubiquitine apparaît relativement dépliée au niveau de l'intermédiaire tandis les résidus 71 et 73 semblent se retrouver dans un environnement relativement non-natif (Figure 4-4). L'approche par sondage global au tryptophane nous a également permis de constater que la conformation de la protéine ne semble pas montrer de signes évidents de la présence de structure de type native durant les premiers millisecondes du repliement (temps mort de l'instrument de mixage) (Figure 4-3B : F_0).

5.1. Mécanisme de repliement d'ubiquitine

Grâce à la méthode d'analyse par sondage global au tryptophane, les travaux de cette thèse ont permis, pour la première fois, de détecter, caractériser, ainsi que la visualiser de manière directe un état intermédiaire transitoire lors du repliement d'ubiquitine (Figure 5-1) après plus de quinze années de recherches effectuées par différents laboratoires dans le monde. Pour mettre en valeur cette découverte, nous présenterons au cours de la prochaine

section un sommaire des différentes données expérimentales pertinentes recueillies à ce jour sur son mécanisme et élaborerons un nouveau modèle de repliement basé sur les résultats de cette thèse (voir figure 1-11 pour l'énumération chronologique des différentes études).⁵

5.1.1. Modèle du repliement d'ubiquitine avant les travaux de cette thèse

Les premières recherches sur le mécanisme de repliement d'ubiquitine ont été initiées au début des années 1990 par l'équipe du Dr Heinrich Roder du Fox Chase Cancer Center à Philadelphie. À l'aide de la méthode d'échange rapide H/D, mise au point dans son laboratoire, il démontra que la majorité des ponts H d'ubiquitine se forment simultanément en moins de 8 ms durant le repliement de la protéine; seuls les ponts H impliquant les résidus 59, 61 et 69, se forment légèrement plus lentement.^{9, pp} Cette même équipe publiera en 1993 la première analyse directe du repliement de la protéine en utilisant la fluorescence du tryptophane inséré à la position 45.¹⁰ Cette étude, ainsi qu'une autre publiée en 1996 par le même laboratoire,¹¹ révéleront la présence d'un signal de fluorescence manquant dans les premiers instants du repliement de la protéine dont l'amplitude semble corrélée avec la vitesse de repliement de la protéine. Ces résultats suggéreront à leurs auteurs qu'ubiquitine se replie via la formation d'un intermédiaire hâtif hydrophobe.¹² À la fin des années 1990, ces résultats représentaient d'ailleurs parmi les évidences les plus citées pour suggérer l'importance des intermédiaires hâtifs hydrophobes dans le repliement des protéines.

^{pp} Étrangement, cette information (la formation légèrement plus lente des ponts-H des résidus 59, 61 et 69 durant le repliement d'ubiquitine) ne sera plus mentionnée dans aucune publication.

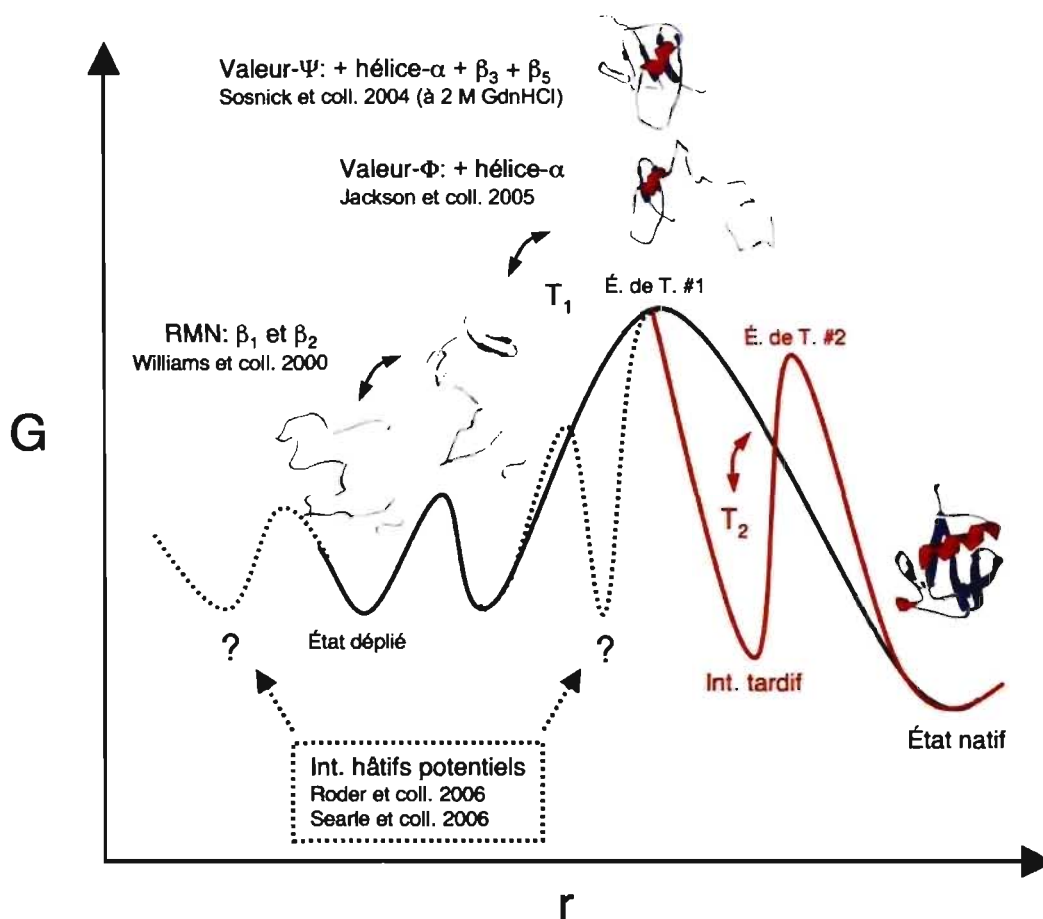


Figure 5-1. Schéma illustrant l'état de nos connaissances sur le mécanisme de repliement de la protéine ubiquitine avant les travaux de cette thèse (ligne noire), et suivant les travaux de cette thèse (ligne rouge). Les lignes pointillées représentent les états intermédiaires hâifs potentiels détectés à l'aide d'un instrument de mixage haute vitesse par l'équipe du professeur Roder (*gauche* : état intermédiaire hors voie; *droite* : état intermédiaire participant au mécanisme de repliement). Des études par RMN menée par l'équipe du Dr Williams ont démontré que les 17 premiers acides aminés d'ubiquitine ont une forte propension à former une épingle de type native en isolation.⁶ Les approches par valeur- Φ ⁷ et valeur- Ψ ⁸ ont apporté deux descriptions relativement différentes de l'état de transition de T₁. Dans le premier cas, comprenant la première épingle et l'hélice- α , dans le deuxième cas, comprenant en plus les brins- β 3 et 5. Toutefois, l'analyse de valeur- Ψ a été réalisée à une concentration de dénaturant à laquelle l'intermédiaire tardif est déstabilisé (2 M GdnHCl). Les travaux présentés dans cette thèse ont quant à eux permis de mettre à jour un intermédiaire tardif dans la voie de repliement d'ubiquitine (voir Figure 5-3 pour des détails sur la structure hypothétique de l'état intermédiaire et sur l'état de transition #2).

À l'automne 2000, l'équipe du Dr Sosnick de l'Université du Chicago publie toutefois une « petite bombe » dans le milieu du repliement des protéines en démontrant que l'amplitude apparemment manquante des premiers instants du repliement de la protéine ubiquitine n'est en fait attribuable qu'à une erreur d'analyse des traces de repliement.¹³ Une fois le temps mort de l'appareil correctement ajusté, et lorsque T_1 n'est pas trop rapide pour être détectée, celle-ci semble posséder toutes les caractéristiques d'une transition de type « deux états ». En 2002, cette équipe publie même une étude réfutant la majorité des évidences expérimentales publiées à ce jour sur les intermédiaires hâtifs de plusieurs protéines.¹⁴ Suivant cette analyse, l'état de transition de la transition de type « deux états » d'ubiquitine (T_1) est alors caractérisé en 2004 par l'équipe du Dr. Sosnick, à l'aide de sa nouvelle méthode d'analyse par valeur- Ψ ,¹⁵ et par l'équipe du Professeur Sophie Jackson de l'Université de Cambridge, à l'aide de la méthode d'analyse par valeur- Φ ¹⁶ (voir section 1.1.2. pour description de ces approches). Ces deux méthodes fourniront toutefois des descriptions significativement différentes de cet état : les valeurs- Φ élevées étant uniquement présentes dans la partie N-terminale d'ubiquitine (1^{ère} épingle et hélice- β) tandis que les valeurs- Ψ incluent en plus la participation du 3^e et du 5^e brin à cet état de transition (Figure 5-1). Les pionniers des deux approches, Sir Alan Fersht de l'Université de Cambridge, et le Dr. Sosnick sentiront même l'obligation de critiquer durement l'approche respective de leur collègue à l'occasion de lettres ouvertes à la revue *Proceedings of the National Academy of Sciences*.^{17, 18}

Entre temps, d'autres évidences concernant les premières transitions potentielles du mécanisme de repliement d'ubiquitine ont également été suggérées par l'analyse de structures résiduelles présentes chez un fragment de cette protéine. L'équipe du Dr Williams du Cambridge Centre for Molecular Recognition a en effet démontré que les 17 premiers acides aminés d'ubiquitine possèdent une forte propension à former une épingle de brin- β de type natif en isolation.⁶

Enfin, tout récemment, de nouvelles évidences provenant des laboratoires du professeur Roder, et du Dr. Searle de l'Université Park à Nottingham, sont venues réalimenter le débat concernant la présence d'états intermédiaires hâtifs durant le repliement d'ubiquitine. D'abord, à l'aide d'un instrument de mixage ultra-rapide,¹⁹ l'équipe du Dr. Roder a mis à jour deux nouvelles transitions plus rapides durant le repliement du mutant Ub^{F45W}.²⁰ Toutefois, aucune évidence ne permet encore de suggérer si ces transitions participent ou non au mécanisme de repliement de la protéine. Finalement, des travaux de l'équipe du Dr Searle ont récemment suggéré, de manière indirecte, la présence d'un intermédiaire hâtif lors du repliement d'ubiquitine puisque les paramètres thermodynamiques issus d'expériences cinétiques pour de nombreux mutants ne correspondent pas à ceux obtenus lors d'expériences d'équilibre (ΔG et valeur de m cinétique plus petites).²¹

5.1.2. Nouveau modèle de repliement d'ubiquitine

En analysant le repliement de la protéine ubiquitine à l'aide de la méthode de sondage par tryptophane, nous avons été en mesure de détecter, caractériser, ainsi que de visualiser pour la première fois un nouvel état intermédiaire tardif dans la voie de repliement de cette protéine (Figure 5-1; ligne rouge). Cet intermédiaire transitoire, formé durant la transition T_1^{qq} affiche une conformation relativement similaire à l'état natif dans sa portion N-terminal (résidus 1 à 54) mais possède toutefois une conformation relativement non-native dans sa portion C-terminal (résidus 57 à 73 : l'hélice-3/10 ainsi que le dernier brin- β). Nos résultats semblent ainsi valider les premières données expérimentales recueillies sur le mécanisme de repliement d'ubiquitine par échange rapide H/D qui suggèrent que les protons amines des résidus 59, 61 et 69 semblent former leur ponts H légèrement plus lentement que le reste des autres protons lors du repliement.

Deux mutants tryptophanes permettent également d'obtenir une description beaucoup plus détaillée de l'état intermédiaire ainsi que du réarrangement I \rightarrow N. Les mutations, L71W et L73W, déstabilisent en effet l'intermédiaire ($\Delta\Delta G_{UI}$) par 5,6 kJ mol⁻¹ et 4,1 kJ mol⁻¹,^π respectivement, sans toutefois affecter la stabilité de l'état natif (ΔG_{UN} = 26,9, 28,2 kJ mol⁻¹ pour les mutants L71W, L73W *versus* 28,2 kJ mol⁻¹ pour la moyenne des mutants) (Figure 5-2B *Gauche*). Il semble logique que la substitution des résidus leucines 71 et 73 par des acides aminés tryptophanes ait peu d'effet sur la stabilité de la forme native car ces positions demeurent relativement exposées au solvant (Figure 5-2A). En revanche, la déstabilisation de l'intermédiaire par ces mêmes mutations implique que ces positions doivent se retrouver enfouies au sein de l'intermédiaire (interactions non-natives). La déstabilisation de l'intermédiaire ne semble pas affecter sa vitesse de formation (T_1) mais accélère plutôt sa vitesse de dépliement ainsi que la vitesse de son réarrangement en l'état natif de manière proportionnelle (Figure 5-2B *Droite*). Ces résultats suggèrent donc que le bris d'interactions non-natives que forment les résidus 71 et 73 au sein de l'intermédiaire fait partie de l'étape limitante du réarrangement (Figure 5-3B). Il est également intéressant de noter que les insertions de tryptophane ayant engendrées les plus fortes réductions de vitesse de T_2 (au moins 2 x) sont retrouvées dans le motif se réarrangeant (N60W, Q62W, et V70W) ou en contact direct avec ce dernier (M1W, L8W, R54W) (Figure 5-2A).

^{ππ} Transition la plus rapide détectée à l'aide d'un instrument de mixage conventionnel et qui fait le plus varier l'intensité de fluorescence du tryptophane 45.

^π La stabilité moyenne de cet intermédiaire étant de $12,9 \pm 2,2$ kJ mol⁻¹.

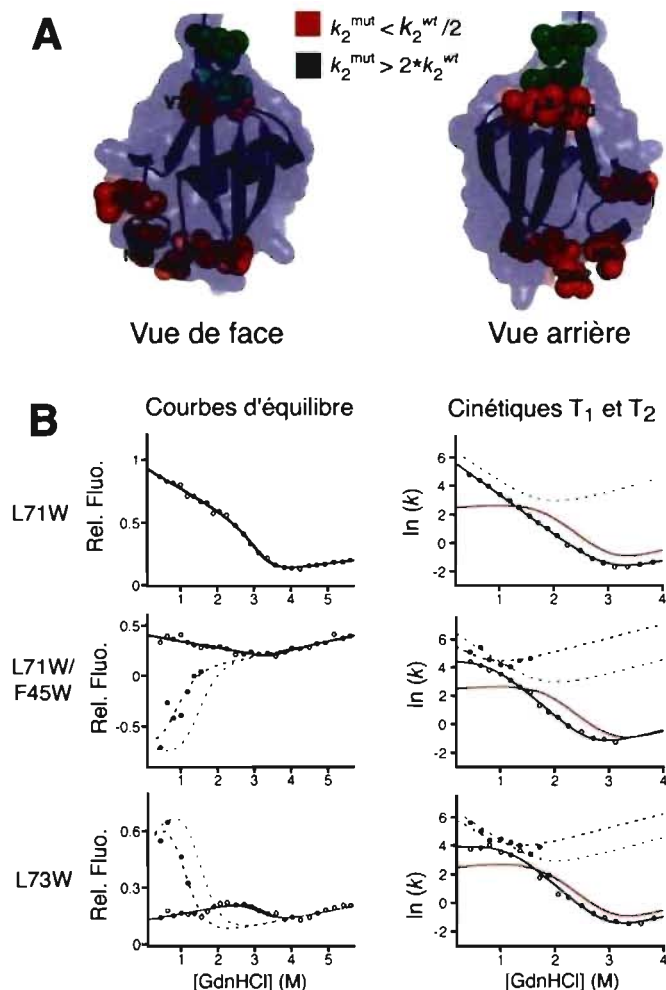


Figure 5-2. Les acides aminés faisant varier la vitesse de T_2 par plus de 2x lorsque mutés en tryptophane sont placés sur le 5^e brin- β ou dans son environnement local (vert : accélération; rouge : ralentissement). A) L'insertion de résidus tryptophanes aux positions 71 et 73 n'affectent pas la stabilité de l'état natif car la chaîne latérale de l'acide aminé pointe en grande partie vers le solvant. B) L'insertion de résidus tryptophanes aux positions 71 et 73 déstabilise de manière sélective l'intermédiaire. *Gauche* : Courbe d'équilibre des états natif (cercles ouverts), et intermédiaires (cercles fermés). Le mutant « double tryptophane » L71W-F45W est utilisé pour détecter l'intermédiaire dans le mutant L71W. Les courbes d'équilibre moyennes des états natif (ligne pleine), et intermédiaire (ligne pointillée), sont indiquées en rouge. *Droite* : La déstabilisation de type non-natives dans l'intermédiaire augmente la vitesse du réarrangement I \rightarrow N de manière proportionnelle (T_2 : cercles vides). Toutefois, la vitesse de repliement de la protéine, T_2 , est rapidement ralentie suite à l'ajout du dénaturant vu la faible stabilité de l'intermédiaire. Les lignes rouges représentent les vitesses moyennes de T_1 (ligne pointillée) et de T_2 (ligne pleine).

Sommaire des paramètres expérimentaux recueillis sur l'état intermédiaire tardif et son réarrangement dans l'état natif.

La somme des évidences obtenues sur l'état intermédiaire ainsi que son réarrangement en état natif rend possible l'élaboration d'un premier modèle de la structure de l'état intermédiaire (Figure 5-3). Bien que ce modèle demeure pour l'instant hypothétique et qualitatif, il s'avérera tout de même d'une grande utilité pour guider les analyses futures sur le mécanisme de repliement de cette protéine.^{ss}

- L'intermédiaire est une forme relativement compacte ($\beta_T \sim 0.86$) et stable ($\sim 12,9 \text{ kJ mol}^{-1}$). De manière générale, la portion N-terminale de la protéine (1-54) affiche une conformation relativement similaire à la forme native (Figure 5-3A). En comparaison, le segment 57-66 semble plutôt déplié. Toutefois, le 5^e brin- β d'ubiquitine n'est sûrement pas complètement exposé au solvant car la majorité des tryptophanes insérés dans cette section, mis à part à la position 66, sont également sensibles à la transition U \rightarrow I. De plus, les mutations L71W et L73W déstabilisent l'état intermédiaire de manière sélective (Figure 5-3B) et l'intensité de fluorescence du tryptophane 73 est beaucoup plus grande dans l'état l'intermédiaire que dans les états natif et déplié (Figure 5-3A). Enfin, la conformation du 5^e brin dans l'état intermédiaire ne doit pas être si éloignée de la conformation native car les tryptophanes insérés dans les brins 1 et 3, qui embrassent le brin 5, adoptent un signal de fluorescence relativement similaire dans les états intermédiaire et natif.

^{ss} Des simulations de types DER "Dynamic Ensemble Refinement"²² réalisées à partir de données expérimentales regroupées au sein de cette thèse permettraient sans doute d'obtenir un modèle structural de l'état intermédiaire plus réaliste et quantitatif.

- Contrairement aux mutations des leucines 73 et 71 qui réduisent la vitesse de repliement d'ubiquitine de manière significative dès l'ajout d'un peu de dénaturant (vu la faible stabilité de l'état intermédiaire pour ces deux mutants –Figure 5-2B), les mutations L69A, et L67A, et L61A ne semblent pas affecter la vitesse de repliement d'ubiquitine même si ces dernières déstabilisent fortement l'état natif (Figure 5-3C –tiré de l'étude de Went & Jackson 2005).⁷ Ces résultats suggèrent ainsi, de manière indirecte, que ces mutations n'affectent pas la stabilité de l'état intermédiaire. De plus, des travaux effectués par l'équipe du Dr Searle semblent suggérer que les mutations T66A et H68A déstabilisent l'état natif de manière spécifique sans affecter l'énergie de l'état intermédiaire car une transition similaire au dépliement de l'intermédiaire tardif est également observée durant des expériences conventionnelles de dépliement de ce mutant.²¹
- L'intermédiaire doit réexposer quelques-uns de ses résidus enfouis (~ 10%) de manière à lui permettre de se réarranger dans en état natif (T_2 accélère avec l'ajout de dénaturant –Figure 3-7). Par exemple, le degré de compaction relatif de l'état intermédiaire est de 0.86 pour le mutant Ub^{M1W} tandis que celui de son état de transition est de 0.79 (Tableau 3-3).
- Le 5^e brin d'ubiquitine comporte quatre résidus leucines espacés à un intervalle de deux acides aminés (i, i+2, i+4, i+6) (L67, L69, L71, et L73). Cette section forme ainsi une longue surface hydrophobe dans une conformation de brin- β .

À partir de toutes ces évidences, nous suggérons donc que les leucines 71 et 73 pourraient très bien remplacer les leucines 67 et 69 pour former le cœur hydrophobe de la protéine au niveau de l'état intermédiaire, sans que ceci ait une grande conséquence pour la stabilité de la protéine (Figure 5-3). Ce décalage du 5^e brin- β vers le bas libérerait ainsi les

leucines 67 et 69 de la structure de la protéine et expliquerait pourquoi la vitesse de repliement d'ubiquitine n'est pas affectée par la mutation de ces résidus.⁷ Cette conformation permettrait d'expliquer pourquoi les mutations T66A et H68A affectent seulement la stabilité de la forme native sans déstabiliser l'intermédiaire tardif car cette portion de la protéine n'est plus fixée en conformation- β au niveau de l'état intermédiaire.^{21, 22} Ce décalage aurait également comme conséquence de laisser quelques résidus en amont de la location 69 exposés au solvant; le segment 57 à 66 s'avère néanmoins relativement hydrophile (Ser⁵⁷-Asp-Tyr-Asp-Ile-Gln-Lys-Glu-Ser-Thr⁶⁶). Cette conformation permettrait également d'expliquer pourquoi les protons amines des résidus 59, 61, et 69, impliqués dans des ponts H au niveau de la structure native sont protégés légèrement plus lentement que le reste des protons de la protéine durant le repliement d'ubiquitine.⁹ Enfin, le réarrangement proposé semble également être conséquent avec l'augmentation marquée de la fluorescence des tryptophanes 1, 12, 42 et 73, observée dans l'état intermédiaire (voir Figure 5-4).

Bien que l'on ait démontré que les résidus L71 et L73 sont en partie responsables pour le lent réarrangement de l'intermédiaire,²³ il se pourrait également que la lenteur de ce réarrangement soit en fait grandement attribuable à la complexité de la structure d'ubiquitine. En effet, l'insertion du 5^e brin au centre du feuillet- β pourrait s'avérer très difficile à réaliser étant donné la grande diversité et quantité de ponts H à former. Par exemple, pour insérer le 5^e brin entre le brin 1 et 3, la structure doit former simultanément

²² Les acides aminés Thr et His ayant une meilleure propension pour la conformation- β que les résidus Ala (Figure 1-5C).

²³ Par exemple, ~ 60% de l'énergie perdue par l'intermédiaire tardif suivant la mutation L71W permet de diminuer la barrière de réarrangement de l'intermédiaire (3,1 / 5,6 kJ mol⁻¹). Voir Figure 5-3A.

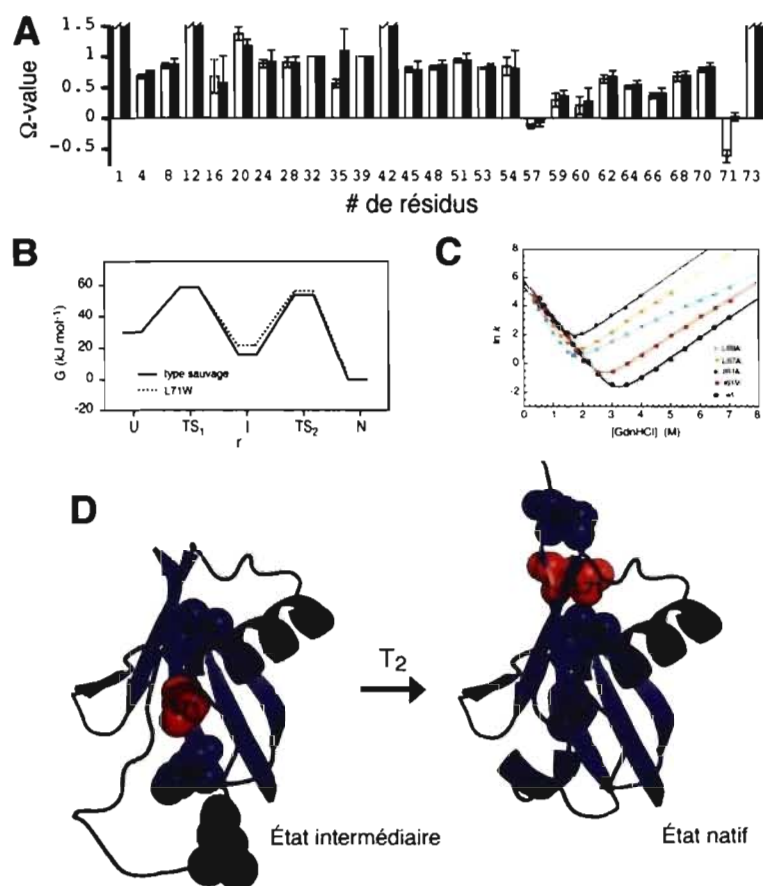


Figure 5-3. Modèle hypothétique de la structure de l'intermédiaire hâtif et de son réarrangement en l'état natif. A) Valeurs- ω obtenues pour l'intermédiaire tardif dans la présente étude (Figure 4-4). La section 57-66 d'ubiquitine apparaît relativement dépliée dans l'intermédiaire tandis les résidus 71 et 73 semblent se retrouver dans un environnement relativement non-natif. B) Les mutations des résidus L73 et surtout L71 en tryptophane déstabilisent spécifiquement l'état intermédiaire, et ralentissent la vitesse de repliement de la protéine sans toutefois affecter l'énergie de l'état natif. Ces mutations permettent également d'accélérer la vitesse du réarrangement de manière proportionnelle. C) Les mutations des leucines 69 et 67 en alanines ont peu d'effet sur la vitesse de repliement malgré le fait qu'elles déstabilisent grandement l'état natif (tiré d'un article de Went & Jackson)²³. D) Structure hypothétique de l'intermédiaire tardif détecté au cours de la présente étude. Un décalage du cinquième brin de quatre résidus vers le bas (les leucines 73 et 71 prenant la place des leucines 69 et 67) dans l'état intermédiaire permet de rationaliser la majorité des évidences expérimentales recueillies sur l'intermédiaire tardif et le réarrangement. Figure effectuée à l'aide de PyMol.

des patrons de ponts H de type parallèle (brin 1) et antiparallèles (brin 3) (voir ces ponts H à la Figure 1-5C). L'implication de cette complexité structurale comme déterminant de la grande barrière d'activation du réarrangement de l'intermédiaire pourrait être en partie étudiée en comparant le mécanisme de protéines possédant des structures similaires à celle d'ubiquitine mais sans identité de séquence apparente. Par exemple, une deuxième transition, sans rapport apparent avec un processus d'isomérisation *cis-trans* de liens peptidiques, a également été détectée durant le repliement du DLR de *raf* (Figure 2-7). Est-ce que cette transition représente également le réarrangement d'un état intermédiaire tardif qui participe au mécanisme de repliement de cette protéine. Si oui, est-ce que l'état intermédiaire et le réarrangement du DLR de *raf* sont similaires à ceux d'ubiquitine. Autant de questions qui pourront trouver réponse dans un futur rapproché grâce à l'approche par sondage au tryptophane présentée dans cette thèse.

Sommaire des évidences recueillies pour l'état hâtif.

L'approche par sondage au tryptophane s'est révélée très efficace pour détecter, caractériser et visualiser un état tardif durant le repliement d'ubiquitine. Toutefois, aucune information directe n'a pu être recueillie sur la présence potentielle d'intermédiaires hâtifs durant le repliement de cette protéine car le temps mort de notre instrument de mixage (3.2 ms) ne permettait pas la détection directe de transitions plus rapide que 350 s^{-1} .^{vv} Il est intéressant toutefois de mentionner que la majorité des tryptophanes ont montré une intensité de fluorescence significativement supérieure à celle de l'état déplié avant la transition T_1 (Figure 4-3C). Cette observation suggère ainsi que la conformation d'ubiquitine avant T_1 pourrait bien être plus compacte que celle de l'état déplié à forte concentration de GdnHCl. En revanche, aucune des positions au sein de la conformation

^{vv} Les phases plus rapides détectées par l'équipe du Dr Roder grâce à un instrument de mixage ultrarapide ont des constances de vitesses de 550 s^{-1} et de 6000 s^{-1} .²⁰

d'ubiquitine pré- T_1 n'a permis de générer une intensité de fluorescence du tryptophane qui soit plus faible que dans l'état déplié comme l'ont faites les positions 16, 35, 45, 57 et 60 dans l'état intermédiaire tardif ou l'état natif (Figure 4-3B). Ce résultat suggère ainsi l'absence apparente de structure de type native stable dans la conformation d'ubiquitine avant la transition T_1 . En revanche, les mutants T12W, R42W, et H68W ont également affiché une déviation négative de leur k_1 à faible concentration de dénaturant (Figure 4-6).^{ww} Cette déviation, souvent interprétée comme un signe indirect de la présence d'un intermédiaire hâtif, a également été observée par l'équipe du Dr Roder lors de leur analyse du repliement d'ubiquitine (mutant F45W) à l'aide d'un instrument de mixage ultra-rapide.²⁰

5.1.3. Expériences à réaliser dans un futur rapproché

À la lumière des résultats présentés et du modèle proposé dans cette thèse, nous sommes maintenant en mesure de proposer plusieurs expériences qui permettraient de valider notre modèle et d'approfondir le mécanisme de repliement d'ubiquitine.

Dans un premier temps, deux expériences simples permettraient de confirmer ou d'infirmer notre modèle qui propose un décalage du 5^e brin dans l'état intermédiaire. Par exemple, si ce décalage est présent, le double mutant L71A-L73A devrait permettre de complètement déstabiliser l'état intermédiaire sans toutefois affecter la stabilité de la forme native. L'analyse du repliement de ce mutant à l'aide de sondes sensibles à l'intermédiaire (par exemple Trp1) devrait permettre de confirmer l'absence de l'état intermédiaire et montrer une diminution proportionnelle de la vitesse de repliement de la protéine. La méthode de valeur- Ψ mise au point par l'équipe du Dr. Sosnick devrait également

^{ww} k_1 devrait normalement varier de manière linéaire en fonction de la concentration de dénaturant.

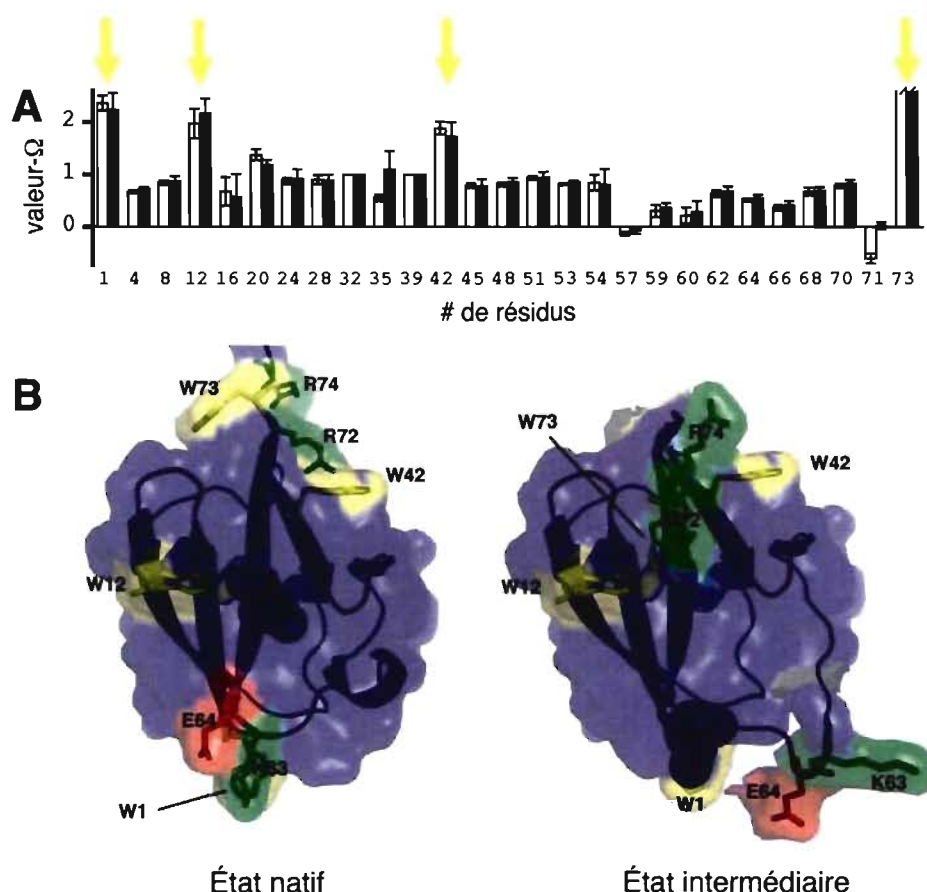


Figure 5-4. Position des tryptophanes présentant une fluorescence non-native dans l'état intermédiaire (fluorescence plus élevée que dans l'état natif). A) Valeur- ω ou fluorescence relative du tryptophane à différentes positions au sein de l'état intermédiaire (0 = similaire à l'état déplié; 1 = de type natif). B) Le modèle d'intermédiaire proposé dans la Figure 5-3 semble conséquent avec le patron de fluorescence non-native obtenu pour les positions 1, 12, 42, et 73. La fluorescence du tryptophane à la position 73 varie peu entre l'état déplié et natif (voir Figure 4-6), ce dernier demeurant majoritairement exposé au solvant, mais augmente grandement au sein de l'état intermédiaire. La conformation proposée pour l'état intermédiaire montre que W73 est complètement enfoui dans le cœur hydrophobe. Le décalage potentiel du 5^e brin dans l'état intermédiaire permet également de rapprocher deux résidus lysines chargés positivement (R72 et R74) à proximité des tryptophanes aux positions 12 et 42 –les charges positives dans le voisinage du tryptophane augmentant généralement l'intensité de fluorescence de ce dernier. Enfin, le tryptophane à la position 1 est relativement exposé au solvant et à proximité de résidus chargés négativement (E64) et positivement (K63) dans l'état natif. Dans l'état intermédiaire, ce dernier se retrouve toutefois en grande partie enfouie sous la leucine 67 (L67).

permettre de confirmer le décalage du 5^e brin via l'étude de divers mutants qui stabiliseraient les brins 5 et 1 ou 5 et 3 dans différents niveaux de décalage. Par exemple, des résidus histidines pourraient être insérés sur chacun des brins à différentes distances l'un de l'autre, et des métaux chargés permettraient alors de stabiliser les deux brins dans plusieurs niveaux de décalage. Les résidus permettant de stabiliser le 5^e brin dans la conformation proposée pour l'intermédiaire devraient ainsi ralentir la vitesse du réarrangement.^{xx}

Une des approches qui permettraient sans doute d'amener le plus de détails structuraux sur l'état intermédiaire tardif serait sans contredit de caractériser par spectroscopie RMN un mutant d'ubiquitine dont l'état intermédiaire serait présent à l'équilibre.²⁴ En effet, comme nous l'avons déjà mentionné, des mutations spécifiques sur une protéine peuvent quelques fois permettre de déstabiliser spécifiquement l'état natif sans toutefois affecter la stabilité d'un de ses états intermédiaires. Par exemple, les évidences présentées dans cette thèse permettent de croire que certaines mutations de résidus dans les segments 57 à 66 d'ubiquitine auraient peu d'impact sur la stabilité de l'état intermédiaire puisque ces résidus semblent adopter une conformation relativement dépliée au sein de cet état. Les mutations T66A et T68A étudiées par l'équipe de Searle semblent d'ailleurs avoir eu ce type d'effet.²¹ À la lumière de nos résultats, le groupe du Dr Sosnick est présentement en train de mettre à profit une telle approche pour caractériser plus en détail l'état intermédiaire tardif détecté dans la présente étude (Tobin Sosnick; communication personnelle).

^{xx} Malheureusement tous les travaux du Dr Sosnick portant sur le repliement d'ubiquitine à l'aide de l'analyse par valeur- Ψ ont été « mystérieusement » réalisés à une concentration de GdnHCl de 2 M là où l'état intermédiaire est déstabilisé... Pourquoi donc? Étrangement, le Dr Sosnick n'a jamais voulu éclaircir avec moi cette question! Je vous laisse donc le soin de tirer vos propres conclusions...

Comme nous l'avons mentionné plus tôt, il serait également très excitant d'analyser les deux premières phases détectées lors du repliement du DLR de raf à l'aide de l'approche par sondage global au tryptophane de manière à déterminer si ces deux protéines, aux structures similaires, se replient également via des intermédiaires tardifs similaires. J'ai d'ailleurs eu « l'occasion » de réaliser ces expériences au cours de mon doctorat,^{yy} mais les résultats de ces études seront sans doute emportés avec moi dans ma tombe vu le manque d'argent et de temps nécessaire à la préparation d'un quatrième manuscrit. Avis tout de même aux intéressés...

Enfin, un des aspects les plus controversés du mécanisme de repliement d'ubiquitine concerne la nature des deux transitions rapides détectées par l'équipe du professeur Roder dans la première milliseconde du repliement d'ubiquitine.²⁰ Le principal problème lié à la caractérisation de ces états demeure le fait que ces derniers ont une durée de vie beaucoup trop courte pour rendre possible des expériences de double saut « repliement-dépliement » qui permettraient de vérifier si ces états participent ou non au mécanisme de repliement (c'est-à-dire s'ils se réarrangent plus vite qu'ils se déplient). En ce sens, l'approche par sondage au tryptophane permettrait sans doute de mettre à jour des sondes de repliement qui soient suffisamment sensibles pour permettent d'obtenir une assez bonne résolution des vitesses de formation et de réarrangement de ces états à différentes concentrations de dénaturant.^{zz}

^{yy} Le repliement du DLR de raf a été analysé à l'aide de 12 mutants tryptophanes.

^{zz} Par exemple, dans la présente étude, la grande sensibilité des sondes W57 et W1 ont permis de déterminer la nature de l'intermédiaire tardif d'ubiquitine sans avoir recours à des expériences de double sauts (Figure 3-5).

5.1.4. Validation des simulations théoriques à partir de modèles simplifiés

Comme nous l'avons mentionné dans l'introduction, les ordinateurs actuels ne possèdent pas encore la puissance nécessaire pour modéliser le repliement de petites protéines en temps réel dans tous ses détails.²⁵ Toutefois, plusieurs laboratoires tentent néanmoins de simuler le repliement de protéines à partir de modèles et d'approches simplifiés.^{26, 27, 28} Bien que ces modèles parviennent quelques fois à prédire l'état de transition d'une protéine, sa vitesse de repliement, ou la présence d'intermédiaire, il demeure toutefois essentiel de continuer de tester leur pouvoir prédictif lorsque de nouvelles évidences expérimentales sont générées.²⁹ Les évidences expérimentales obtenues sur l'intermédiaire tardif au sein de la présente étude représentent donc une excellente occasion pour valider le réalisme de certaines simulations réalisées sur le repliement d'ubiquitine. À cet effet, deux simulations retiennent particulièrement notre attention. L'équipe du Dr Wand de l'Université de Nanjing en Chine a simulé le repliement d'ubiquitine en utilisant des potentiels de type « Go » qui tiennent juste compte des interactions de type natives. Ces travaux ont permis d'identifier une voie de repliement parallèle mineure impliquant un intermédiaire mal replié fortement stabilisé par des interactions entre le brin 1 et 5 de la protéine;³⁰ le réarrangement de cet intermédiaire se produisant via le bris des interactions entre ces deux brins. Des simulations effectuées par l'équipe de Theresa Head-Gordon de l'Université de Californie à Berkeley à partir d'une représentation simplifiée d'ubiquitine utilisant trois types d'acide aminés (hydrophobe, hydrophile, et neutre) ont également permis de mettre à jour un intermédiaire tardif qui semble en tout point répondre aux contraintes structurales obtenues au sein de cette thèse.³¹ Dans cette simulation, la queue C-terminale d'ubiquitine forme ses interactions natives de manière non-coopérative à la toute fin du repliement. Dans cet intermédiaire, la région impliquant le dernier tour est alors repliée en hélice tandis que le 5^e brin vient se coller au feuillet- β sans toutefois être inséré dans la structure. Le fait que deux modèles simplifiés, complètement distincts, arrivent tous deux à mettre à jour un problème de repliement

impliquant le 5^e brin d'ubiquitine vient renforcer l'hypothèse voulant que l'intermédiaire tardif serait en grande partie attribuable à la complexité de la structure d'ubiquitine.^{aaa} Ces résultats tendent également à suggérer que les simulations théoriques de repliement à partir de modèles simplifiés semblent fournir une description assez réaliste du repliement des protéines.

5.2. Avantages, limitations, et perspectives de la méthode de sondage des structures protéiques par tryptophane

Au cours de la prochaine section, nous présenterons un sommaire des avantages et des limitations inhérentes à l'utilisation de l'approche de sondage par tryptophane pour détecter, caractériser, et visualiser les structures transitoires de protéines. En guise de perspective, nous proposerons ensuite quelques applications potentielles de cette approche dans des domaines autres que celui de l'étude du mécanisme de repliement des protéines.

5.2.1. Détection et caractérisation des structures transitoires de protéines

L'utilisation de plusieurs tryptophanes, localisés à la surface d'ubiquitine, pour analyser son mécanisme de repliement s'est révélé très efficace car :

- 1) Tous les sites d'insertion essayés à la surface d'ubiquitine (27) ont permis de générer suffisamment de changement d'intensité de fluorescence du tryptophane pour nous permettre de suivre le repliement/dépliement de la protéine (Figure 4-5);
- 2) ~ 85% des positions (23/27) ont permis au tryptophane inséré de détecter les transitions liées à la formation (T₁) au réarrangement (T₂) de l'intermédiaire tardif

^{aaa} En plus, un deux modèle ne tient même pas compte des interactions non-natives.

(Figure 4-6). Toutefois, seuls cinq de ces positions (1, 12, 42, 57 et 73) ont permis de détecter l'intermédiaire hors de tout doute. Dans les cas où seule une transition était détectée l'ajout d'un tryptophane additionnel, ou l'utilisation de sonde indépendante comme l'ANS (Figure 3-4) a facilement permis de mettre à jour la transition manquante. Par exemple, le double mutant T66W-F45W a été utilisé pour mettre à jour T_1 lors du repliement du mutant T66W (Figure 3-6) et le double mutant L71W-F45W a été utilisé pour mettre à jour l'intermédiaire lors du repliement de L71W (Figure 5-2B).

3) L'insertion de tryptophanes à la surface d'ubiquitine a eu très peu d'effet sur la stabilité de l'état natif (Figure 4-3A) et seuls 3 mutants ont vu la vitesse de formation ou de réarrangement de leur état intermédiaire (ou sa stabilité) varier de manière significative (Figure 4-3B). Cette variation a tout de même permis d'en apprendre davantage sur l'implication énergétique de ces résidus dans le mécanisme de repliement de la protéine (Figure 5-3).

La combinaison de tous ces avantages a ainsi permis de mettre à jour le fait que T_1 ne représente pas une transition de type « deux états » pour ubiquitine, mais bien la formation d'un intermédiaire tardif, et que T_2 ne représente pas une fraction des protéines se repliant plus lentement mais bien le réarrangement de cet intermédiaire en l'état natif. Pour arriver à une telle conclusion, il aura toutefois fallu s'assurer que l'insertion des tryptophanes ne modifie pas le mécanisme de repliement d'ubiquitine, ou ne réduise pas le niveau d'accumulation de l'intermédiaire.^{bbb} Dans la présente étude, ce contrôle a été simplement réalisé en comparant les traces de repliement de tous les mutants à l'aide de la fluorescence d'une tierce molécule, l'ANS, qui est connu pour lier les surfaces

^{bbb} La caractérisation des intermédiaires transitoires peu stables devrait donc nécessiter plus d'expériences contrôle puisqu'une variation d'amplitude d'une transition chez un mutant tryptophane spécifique pourrait simplement être attribuable à la stabilisation ou la déstabilisation de l'intermédiaire suite à la mutation.

hydrophobes des protéines (Figure 3-4).³² Finalement, il est également important de mentionner que pour pouvoir discerner correctement deux transitions qui ont des amplitudes du même signe, il est nécessaire que leurs vitesses soient séparées par un facteur d'au moins trois. Dans la présente étude, la vitesse de T_1 a d'ailleurs été accélérée sélectivement par rapport à celle de T_2 via une augmentation de la température (30 °C) de manière à permettre une meilleure résolution des deux transitions.

Finalement, l'approche par sondage global au tryptophane devrait représenter un outil complémentaire indispensable aux approches par valeur- Φ et par valeur- Ψ car pour être efficaces, ces techniques doivent avant tout pouvoir compter sur des sondes permettant de détecter efficacement toutes les transitions de repliement.

5.2.1. Visualisation des structures transitoires de protéines

On a vu à la Figure 1-8C que l'approche par valeur- Φ demeure grandement limitée pour caractériser les structures d'intermédiaires transitoires puisque les mutations utilisées dans ce type d'analyse perturbent grandement la stabilité de la protéine et la vitesse des transitions détectées.^{ccc} Une des conséquences de ces larges déstabilisations est le fait qu'il devient alors très difficile d'obtenir de l'information concernant la stabilité d'un intermédiaire si la vitesse de formation de l'intermédiaire et celle de son réarrangement deviennent similaires; l'intermédiaire devenant alors indiscernable dans les courbes de repliement (Figure 1-8C). Au contraire, l'insertion de résidus tryptophanes à la surface d'une protéine semble généralement peu perturbante (Figure 4-2A); les transitions de formation et de réarrangement de l'intermédiaire peuvent alors être beaucoup plus facilement identifiées.

^{ccc} Sir Alan Fersht, pionnier de l'approche par valeur- Φ , recommande habituellement des déstabilisations supérieures à 5 kJ mol⁻¹.¹⁶

L'approche par la valeur- ω ^{ddd} introduite dans la présente thèse permet de déterminer les régions de la protéine qui possèdent une conformation native au niveau de l'intermédiaire et les positions qui se retrouvent plutôt dans un environnement de type déplié ou non-natif. De plus, les insertions de tryptophanes qui entraînent une déstabilisation de l'intermédiaire, ou une variation significative des vitesses de formation ou réarrangement de l'intermédiaire, permettent également de fournir davantage d'informations structurales. Par exemple, les mutations L71W et L73W, présentées à la Figure 5-2, démontrent que le 5^e brin d'ubiquitine forme des interactions de type non-natives au niveau de l'intermédiaire. En ce sens, les différents mutants tryptophanes créés à l'occasion d'une analyse de valeur- ω s'avèrent grandement complémentaires à l'approche de valeur- Φ car ces mutations sont réalisées à la surface de la protéine; les mutations déstabilisantes utilisées pour établir des valeur- Φ ciblant généralement des résidus enfouis, impliqués dans plusieurs interactions natives.^{16, eee}

Toutefois, pour que l'approche par valeur- ω demeure simple et efficace, il est souhaitable que l'intermédiaire transitoire à l'étude soit relativement stable par rapport à l'état déplié de manière à éviter que son niveau d'accumulation ne diminue dépendamment de la position du tryptophane inséré. En ce sens, la stabilité de l'intermédiaire tardif d'ubiquitine était idéale (ΔG_{UI} moyen = 12,9 kJ mol⁻¹) puisque des mutations déstabilisant cet intermédiaire par plus de 5 kJ mol⁻¹ (comme la mutation L71W –Figure 5-2) permettaient encore à plus de 95% de ce dernier d'être accumulé après T₁. La valeur de fluorescence d'un tryptophane dans l'état intermédiaire transitoire serait relativement

^{ddd} Nous avons décidé de baptiser notre approche « analyse par valeur- ω » car cette lettre grecque rappelle le « W » utilisé pour représenter l'acide aminé tryptophane.

^{eee} Les mutations des résidus de surfaces 71 et 73 n'auraient jamais été mutés dans une analyse par valeur- Φ .

difficile à évaluer pour plusieurs mutants si l'on devait en plus estimer pour chaque mutant la fraction de l'intermédiaire s'accumulant.

Un autre problème rencontré lors de l'étude de mutants tryptophanes peut être lié au fait que certains tryptophanes n'arrivent pas à détecter l'une, ou l'autre, des transitions structurales.^{fff} Cette situation peut survenir lorsque la fluorescence de l'intermédiaire est la même que celle de l'état déplié (T66W), ou est identique à celle de l'état natif (D32W et D39W). Toutefois, il est également possible que la non détection d'une transition soit attribuable au fait que les vitesses de formation et de réarrangement de l'intermédiaire soient trop similaires pour permettre leur résolution (L71W –Table 4-1). En revanche, la stratégie du double mutant tryptophane permet de résoudre cette ambiguïté en confirmant la présence des transitions manquantes et en permettant également d'estimer de manière indirecte la valeur de fluorescence d'un tryptophane à une position précise dans l'état intermédiaire à partir des valeurs obtenues pour le double mutant et de l'autre simple mutant. Par exemple, le double mutant F45W-T66W a permis de confirmer la présence de T_1 durant le repliement du mutant T66W (Figure: 3-6). L'intensité de fluorescence du Trp 66 au niveau de l'intermédiaire tardif a également pu être correctement estimée à partir des intensités de fluorescence de l'intermédiaire des mutants F45W-T66W et F45W.^{ggg} Aussi, les transitions T_1 et T_2 du mutant L71W ont pu être seulement détectées à l'aide du double mutant F45W-L71W; ce dernier a également permis d'évaluer, de manière indirecte, la fluorescence de Trp 71 dans l'intermédiaire à partir des intensités de fluorescence de l'intermédiaire des mutant F45W-L71W et F45W.^{hhh}

^{fff} Quatre mutants tryptophanes sur 27 dans notre étude ne détectaient qu'une seule transition (D32W, D39W, T66W, et L71W).

$$^{ggg} F_I^{T66W} = F_I^{F45W-T66W} - F_I^{F45W}$$

$$^{hhh} F_I^{L71W} = F_I^{F45W-L71W} - F_I^{F45W}$$

Une des avenues qui reste à explorer consiste à déterminer si les valeur- ω peuvent servir de contraintes quantitatives lors des simulations de DER?²² La question consiste en fait à déterminer quelle interprétation structurale donner à cette valeur. On a vu que l'analyse par valeur- ω peut s'avérer très utile pour suggérer la structure d'un état de manière qualitative et pour guider l'ingénieur de protéine dans ses choix de mutations à effectuer pour stabiliser sélectivement un état intermédiaire transitoire. Toutefois, ce paramètre pourrait-il servir de contrainte réaliste pour mieux simuler la structure d'un état intermédiaire? La problématique liée à l'interprétation des différents paramètres expérimentaux issus d'approches diverses ne date pas d'hier. Par exemple, on assume généralement que la valeur- Φ est équivalente à la fraction des contacts de type natif dans un état de transition.¹⁶ Toutefois, cette définition demeure largement controversée lorsque cette valeur- Φ est partielle, inférieure à 0, ou supérieure à 1.^{18, 33, 34} Néanmoins, à partir de cette simple définition, des simulations DER de l'état de transition de quelques protéines ont été réalisées et se sont avérées d'une remarquable reproductibilité (voir résultat pour domaine SH3 à la Figure 1-7).²² D'une manière similaire, des simulations DER d'un état intermédiaire ont également été réalisées en utilisant le déplacement chimique relatif de ses protons, détectés lors d'expériences RMN, comme paramètre structural suggérant une conformation native ou dépliée.³⁵ Nous suggérons donc que la valeur- ω pourrait représenter la fraction de contact de type natif présent au niveau de chaque résidu de l'état intermédiaire? Cette hypothèse devra cependant être explorée dans un avenir rapproché via des modélisations faites à partir de cette variable.

Dans un futur rapproché, l'avancement des recherches sur la modélisation de l'intensité de fluorescence du tryptophane, en fonction de sa location dans une protéine (Figure 1-9D),³⁶ contribuera grandement à populariser l'approche par valeur- ω en la rendant plus prédictive et donc, plus puissante. Cette connaissance pourrait nous permettre, en autre chose, d'établir des contraintes structurales plus spécifiques de l'environnement chimique local du tryptophane dans un état transitoire à partir de l'intensité de son signal de

fluorescence. D'autres caractéristiques de fluorescence du tryptophane, comme ses temps de déclin d'anisotropie ou de son intensité de fluorescence en picoseconde/nanoseconde, pourraient également fournir des contraintes structurales beaucoup plus spécifique à différentes locations dans la protéine.^{37, iii} Des expériences de FRET mettant en vedette les différents tryptophanes comme donneur et un accepteur pourraient également servir à établir de nombreuses contraintes de distance entre différentes régions de la protéine dans un état donné.³⁸ Ces données expérimentales permettraient ainsi de fixer des contraintes structurales plus spécifiques résultant ainsi dans une meilleure description de l'état intermédiaire.

5.2.3 Autres applications envisagées de l'approche par sondage global au tryptophane

L'approche par sondage global au tryptophane devrait également être applicable à l'étude d'intermédiaires transitoires détectés lors de mécanismes de fonctionnement des protéines. Par exemple, cette approche pourrait servir à caractériser les changements structuraux prenant place au sein d'une protéine durant une réaction enzymatique ou lorsque celle-ci se lie à une autre protéine, à une drogue ou un inhibiteur. L'analyse d'une fonction de protéine à l'aide de plusieurs mutants tryptophanes permettrait autant de caractériser la cinétique des différentes transitions détectées que de caractériser les changements conformationnels prenant place durant ces transitions. Cette approche permettrait en plus d'obtenir de nombreuses contraintes expérimentales advenant le fait que l'insertion du tryptophane annule la fonction ou modifie grandement sa dynamique. Finalement, l'approche par valeur- ω guiderait l'ingénieur de protéine dans son choix de mutations permettant de déstabiliser sélectivement l'état natif pour accroître la population l'état

ⁱⁱⁱ Tel que le degré d'hétérogénéité du polypeptide.

transitoire à l'équilibre.³⁹ La caractérisation de l'état transitoire pourrait alors être menée de manière plus détaillée à l'aide de la spectroscopie RMN.²⁴

L'approche par sondage au tryptophane pourrait également trouver de nombreuses applications dans l'industrie pharmaceutique en permettant notamment de réaliser des essais de criblage de drogues à grande échelle. Par exemple, une librairie de drogues pourrait facilement être testée sur des plaques de 96 puits contenant différents mutants tryptophanes d'une protéine d'intérêt. Une drogue qui lierait sélectivement la protéine serait alors détectée par plusieurs tryptophanes qui afficheraient une variation de fluorescence significative. Cette stratégie s'avérerait particulièrement efficace pour mettre à jour de nouveaux sites allostériques très prisés par l'industrie pharmaceutique comme sites potentiels de liaison de drogues.⁴⁰ Les mutants tryptophanes affichant le plus grand changement de fluorescence suivant la liaison à la drogue pourraient alors être facilement utilisés pour caractériser le mécanisme d'action de la drogue.

Enfin, une variation de la stratégie par sondage au tryptophane pourrait également être appliquée à l'étude des changements conformationnels d'une protéine dans son contexte cellulaire, c'est-à-dire *in vivo*. En effet, les mécanismes d'un très grand nombre de protéines ne pourront jamais être élucidés *in vitro* puisque leur fonction nécessite la présence d'un contexte cellulaire extrêmement complexe (par exemple, les récepteurs membranaires ou les protéines participant à des megacomplexes). La fluorescence de l'acide aminé tryptophane est inutile pour des études *in vivo* car la cellule fluoresce trop à ces longueurs d'ondes. Toutefois, les techniques modernes de clonages rendent maintenant possible l'insertion sélective d'acides aminés modifiés au sein d'une protéine.⁴¹ Certains de ces résidus peuvent même émettre un signal de fluorescence se démarquant du « bruit » de la cellule.^{42, 43} Une approche de sondage global à l'aide de tels résidus rendrait ainsi possible la détection de changements conformationnels très subtils dans le temps et dans l'espace à l'intérieur de la cellule et offrirait un monde de possibilités pour le design de nouveaux biosenseurs.^{44, 45}

5.3. Références

1. Plaxco, K. W., Simons, K. T., Ruczinski, I. & Baker, D. (2000). Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics. *Biochemistry* **39**, 11177-83.
2. Campbell-Valois, F. X. & Michnick, S. W. (2007). The transition state of the ras binding domain of Raf is structurally polarized based on Phi-values but is energetically diffuse. *J Mol Biol* **365**, 1559-77.
3. Pappenberger, G., Aygun, H., Engels, J. W., Reimer, U., Fischer, G. & Kiefhaber, T. (2001). Nonprolyl cis peptide bonds in unfolded proteins cause complex folding kinetics. *Nat Struct Biol* **8**, 452-8.
4. Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* **14**, 4953-63.
5. Jackson, S. E. (2006). Ubiquitin: a small protein folding paradigm. *Org Biomol Chem* **4**, 1845-53.
6. Zerella, R., Chen, P. Y., Evans, P. A., Raine, A. & Williams, D. H. (2000). Structural characterization of a mutant peptide derived from ubiquitin: implications for protein folding. *Protein Sci* **9**, 2142-50.
7. Went, H. M. & Jackson, S. E. (2005). Ubiquitin folds through a highly polarized transition state. *Protein Eng Des Sel* **18**, 229-37.
8. Krantz, B. A., Dothager, R. S. & Sosnick, T. R. (2004). Discerning the structure and energy of multiple transition states in protein folding using psi-analysis. *J Mol Biol* **337**, 463-75.
9. Briggs, M. S. & Roder, H. (1992). Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci U S A* **89**, 2017-21.
10. Khorasanizadeh, S., Peters, I. D., Butt, T. R. & Roder, H. (1993). Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* **32**, 7054-63.
11. Khorasanizadeh, S., Peters, I. D. & Roder, H. (1996). Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol* **3**, 193-205.
12. Roder, H. & Colon, W. (1997). Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol* **7**, 15-28.
13. Krantz, B. A. & Sosnick, T. R. (2000). Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**, 11696-701.

14. Krantz, B. A., Mayne, L., Rumbley, J., Englander, S. W. & Sosnick, T. R. (2002). Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J Mol Biol* **324**, 359-71.
15. Sosnick, T. R., Krantz, B. A., Dothager, R. S. & Baxa, M. (2006). Characterizing the protein folding transition state using psi analysis. *Chem Rev* **106**, 1862-76.
16. Fersht, A. R. (1999). *Structure and mechanism in protein science : a guide to enzyme catalysis and protein folding*, Freeman, W.H., New-York.
17. Fersht, A. R. (2004). Phi value versus psi analysis. *Proc Natl Acad Sci U S A* **101**, 17327-8.
18. Sosnick, T. R., Dothager, R. S. & Krantz, B. A. (2004). Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. *Proc Natl Acad Sci U S A* **101**, 17377-82.
19. Roder, H., Maki, K., Cheng, H. & Shastry, M. C. (2004). Rapid mixing methods for exploring the kinetics of protein folding. *Methods* **34**, 15-27.
20. Roder, H., Maki, K. & Cheng, H. (2006). Early events in protein folding explored by rapid mixing methods. *Chem Rev* **106**, 1836-61.
21. Crespo, M. D., Simpson, E. R. & Searle, M. S. (2006). Population of on-pathway intermediates in the folding of ubiquitin. *J Mol Biol* **360**, 1053-66.
22. Vendruscolo, M. & Dobson, C. M. (2005). Towards complete descriptions of the free-energy landscapes of proteins. *Philos Transact A Math Phys Eng Sci* **363**, 433-50; discussion 450-2.
23. Went, H. M., Benitez-Cardoza, C. G. & Jackson, S. E. (2004). Is an intermediate state populated on the folding pathway of ubiquitin? *FEBS Lett* **567**, 333-8.
24. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Di Nardo, A. A., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004). Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430**, 586-90.
25. Wolynes, P. G. (2005). Energy landscapes and solved protein-folding problems. *Philos Transact A Math Phys Eng Sci* **363**, 453-64; discussion 464-7.
26. Shea, J. E. & Brooks, C. L., 3rd. (2001). From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu Rev Phys Chem* **52**, 499-535.
27. Mirny, L. & Shakhnovich, E. (2001). Protein folding theory: from lattice to all-atom models. *Annu Rev Biophys Biomol Struct* **30**, 361-96.
28. Scheraga, H. A., Khalili, M. & Liwo, A. (2007). Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem* **58**, 57-83.

29. Alm, E. & Baker, D. (1999). Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc Natl Acad Sci U S A* **96**, 11305-10.
30. Zhang, J., Qin, M. & Wang, W. (2005). Multiple folding mechanisms of protein ubiquitin. *Proteins* **59**, 565-79.
31. Sorenson, J. M. & Head-Gordon, T. (2002). Toward minimalist models of larger proteins: a ubiquitin-like protein. *Proteins* **46**, 368-79.
32. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F. & Gilmanshin, R. I. (1991). Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-28.
33. Sanchez, I. E. & Kiefhaber, T. (2003). Origin of unusual phi-values in protein folding: evidence against specific nucleation sites. *J Mol Biol* **334**, 1077-85.
34. Weikl, T. R. & Dill, K. A. (2007). Transition-states in protein folding kinetics: the structural interpretation of Phi values. *J Mol Biol* **365**, 1578-86.
35. Lindorff-Larsen, K., Kristjansdottir, S., Teilum, K., Fieber, W., Dobson, C. M., Poulsen, F. M. & Vendruscolo, M. (2004). Determination of an ensemble of structures representing the denatured state of the bovine acyl-coenzyme A binding protein. *J Am Chem Soc* **126**, 3291-9.
36. Callis, P. R. & Liu, T. (2004). Quantitative prediction of fluorescence quantum yields for tryptophan in proteins. *J Phys Chem B* **108**, 4248-4259.
37. Beechem, J. M. (1997). Picosecond fluorescence decay curves collected on millisecond time scale: direct measurement of hydrodynamic radii, local/global mobility, and intramolecular distances during protein-folding reactions. *Methods Enzymol* **278**, 24-49.
38. Magg, C., Kubelka, J., Holtermann, G., Haas, E. & Schmid, F. X. (2006). Specificity of the initial collapse in the folding of the cold shock protein. *J Mol Biol* **360**, 1067-80.
39. Sanz, J. M. & Fersht, A. R. (1993). Rationally designing the accumulation of a folding intermediate of barnase by protein engineering. *Biochemistry* **32**, 13584-92.
40. Hardy, J. A. & Wells, J. A. (2004). Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* **14**, 706-15.
41. Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z. & Schultz, P. G. (2003). An expanded eukaryotic genetic code. *Science* **301**, 964-7.
42. Wang, J., Xie, J. & Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *J Am Chem Soc* **128**, 8738-9.

43. Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J. W. & Schultz, P. G. (2006). A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **103**, 9785-9.
44. Wilson, G. S. & Gifford, R. (2005). Biosensors for real-time in vivo measurements. *Biosens Bioelectron* **20**, 2388-403.
45. Giepmans, B. N., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. (2006). The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-24.

Annexe

Manifeste de l'International Foldiomics Consortium (IFC) dont l'objectif vise à établir des normes standardes pour l'étude du mécanisme de repliement des protéines à la communauté scientifique.

Protein folding: Defining a “standard” set of experimental conditions and a preliminary kinetic data set of two-state proteins

KAREN L. MAXWELL,¹ DAVID WILDES,⁵ ARASH ZARRINE-AFSAR,² MIGUEL A. DE LOS RIOS,⁶ ANDREW G. BROWN,⁷ CLAIRE T. FRIEL,⁸ LINDA HEDBERG,⁹ JIA-CHERNG HORNG,¹⁰ DIANE BONA,¹ ERIK J. MILLER,⁵ ALEXIS VALLÉE-BÉLISLE,¹¹ EWAN R.G. MAIN,⁷ FRANCESCO BEMPORAD,¹² LINLIN QIU,¹³ KAARE TEILUM,¹⁴ NGOC-DIEP VU,¹⁵ ALED M. EDWARDS,^{1,3,4} INGO RUCZINSKI,¹⁶ FLEMMING M. POULSEN,¹⁴ BIRTHE B. KRAGELUND,¹⁴ STEPHEN W. MICHNICK,¹¹ FABRIZIO CHITI,¹² YAWEN BAI,¹⁵ STEPHEN J. HAGEN,¹³ LUIS SERRANO,¹⁷ MIKAEL OLIVEBERG,⁹ DANIEL P. RALEIGH,¹⁰ PERNILLA WITTUNG-STAFSHED,¹⁸ SHEENA E. RADFORD,⁸ SOPHIE E. JACKSON,⁷ TOBIN R. SOSNICK,¹⁹ SUSAN MARQUESE,⁵ ALAN R. DAVIDSON,^{2,4} AND KEVIN W. PLAXCO⁶

¹Ontario Cancer Institute and Department of Medical Biophysics, ²Department of Biochemistry, ³Structural Genomics Consortium, and ⁴Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada

⁵Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA

⁶Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, California 93106, USA

⁷Laboratory of Biochemistry, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

⁸School of Biochemistry and Microbiology and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

⁹Department of Biochemistry, Umea University, S-901 87 Umea, Sweden

¹⁰Department of Chemistry, State University of New York, Stony Brook, New York 11794, USA

¹¹Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada

¹²Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, 50134 Firenze, Italy

¹³Department of Physics, University of Florida, Gainesville, Florida 32611, USA

¹⁴Institute of Molecular Biology, Department of Protein Chemistry, DK-1353 Copenhagen, Denmark

¹⁵Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

¹⁶Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, USA

¹⁷European Molecular Biology Laboratory, Heidelberg D-69012 Germany

¹⁸Biochemistry and Cell Biology Department and Chemistry Department, Rice University, Houston, Texas 77251, USA

¹⁹Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637, USA

(RECEIVED October 28, 2004; FINAL REVISION November 26, 2004; ACCEPTED November 26, 2004)

Abstract

Recent years have seen the publication of both empirical and theoretical relationships predicting the rates with which proteins fold. Our ability to test and refine these relationships has been limited, however, by a

Reprint requests to: Kevin W. Plaxco, Department of Chemistry and Biochemistry, University of California, Santa Barbara, Santa Barbara, CA 93106, USA; e-mail: [REDACTED]

Abbreviations: GuHCl, guanidine hydrochloride; tris, tris hydroxymeth-

ylaminoethane; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; TCEP, tris(2-carboxyethyl)phosphine; CD, circular dichroism.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.041205405>.

variety of difficulties associated with the comparison of folding and unfolding rates, thermodynamics, and structure across diverse sets of proteins. These difficulties include the wide, potentially confounding range of experimental conditions and methods employed to date and the difficulty of obtaining correct and complete sequence and structural details for the characterized constructs. The lack of a single approach to data analysis and error estimation, or even of a common set of units and reporting standards, further hinders comparative studies of folding. In an effort to overcome these problems, we define here a "consensus" set of experimental conditions (25°C at pH 7.0, 50 mM buffer), data analysis methods, and data reporting standards that we hope will provide a benchmark for experimental studies. We take the first step in this initiative by describing the folding kinetics of 30 apparently two-state proteins or protein domains under the consensus conditions. The goal of our efforts is to set uniform standards for the experimental community and to initiate an accumulating, self-consistent data set that will aid ongoing efforts to understand the folding process.

Keywords: two-state; protein folding; kinetics; chevron plots; equilibrium

Supplemental material: see www.proteinscience.org

The folding kinetics of dozens of proteins have been reported to date (e.g., Jackson 1998). The conditions under which these data were obtained, however, vary widely in terms of temperature (e.g., reference temperatures ranging from 10°C–41°C) (Ferguson et al. 1999; Spector and Raleigh 1999), pH, ionic strength, and data reduction methods. Similarly, the relevant raw kinetic data are rarely made available, and the relevant structural data for the characterized construct are often difficult to obtain unambiguously. As a consequence, no systematic, fully detailed database is available with which to define and develop quantitative, testable models of folding kinetics. In an effort to overcome this problem, we describe here a set of experimental conditions, data analysis methods, and reporting standards that we hope will provide a benchmark for the experimental folding community. We argue strongly for the adoption of a standard reference temperature of 25°C, and on the prominent, clear, and complete reporting of other conditions (pH, denaturant, experimental approach, etc.). We place equal importance on the availability of raw kinetic data and on providing the sequence and structural information required for a complete description of the characterized construct.

In the following sections, we describe "consensus" conditions, fitting algorithms, error estimation methods, and structural data that we believe represent the minimum complete description of any new experimental folding system. Lastly, we describe the folding of some 30 apparently two-state single-domain proteins or isolated protein domains, the folding kinetics of seven of which have not previously been reported, under the herein recommended conditions.

Standard conditions

Protein folding rates are sensitive to a wide variety of environmental conditions, including temperature, pH, buffer, ionic strength, and the concentration and nature of any re-

sidual denaturant. Here we propose standard values for each of these parameters, and we encourage researchers in the field to use these values *whenever possible*. Naturally, if for some proteins the adoption of one of the standards significantly complicates the experimentalist's task, departure from these conditions is reasonable. Excluding temperature, for which we believe that reporting values at the consensus temperature is very strongly preferred, we believe it is more important to report the conditions employed as prominently and fully as the results themselves than to adhere to these necessarily somewhat arbitrary standards. Instead we offer these consensus conditions in the hope that they will be employed whenever possible, so as to reduce the number of variables contributing to the available kinetic data and thus simplify the comparison of folding data obtained by different laboratories.

Temperature

Protein folding rates display a strong dependence on temperature. Folding reactions commonly exhibit activation enthalpies of 10–20 kJ/mol (e.g., Jackson and Fersht 1991b; Plaxco et al. 1998a; Main et al. 1999), corresponding to a temperature sensitivity of ~1.5%–3% per degree Celsius. We thus strongly encourage researchers in the field to report the folding kinetics observed at a single consensus temperature. Arguments can be made for consensus temperature as low as 10°C, where rates are slower and there are fewer artifacts associated with cavitation and degassing, and as high as 37°C, which is at or near the physiological temperature of mammals, birds, and many experimentally important prokaryotes. The present investigators have agreed on a compromise experimental temperature of 25°C based on three arguments: First, 25°C is slightly above room temperature and thus easily maintained via thermostatically controlled heating. Second, these conditions are far enough above the freezing point that they are an acceptable end-

point for most temperature-jump approaches. Third, 25°C is the most commonly employed temperature in the previous folding literature and is near the second most commonly employed temperature, 22°C; the forward adoption of this standard thus maximizes backward compatibility with the literature. We thus strongly recommend that folding kinetic studies report rates measured at 25°C whenever possible.

Chemical denaturants

In the majority of folding kinetics experiments, chemical denaturants—typically urea or guanidinium salts—are employed to destabilize the native state. Rates are then measured in the presence of varying amounts of these denaturants. Because ionic strength is a potentially confounding variable and linear extrapolation is generally less applicable to guanidinium salts (Makhatadze 1999), we recommend urea as a denaturant. However, since many small proteins do not unfold in saturated urea solutions at 25°C, an alternative denaturant, such as guanidinium chloride or guanidinium thiocyanate, may be required. Where the use of guanidinium salts is known to complicate the analysis of folding kinetics because of ionic strength effects, simultaneous pH- and urea concentration-jump techniques should be employed (Went et al. 2004). The precise mode of denaturation should be identified as prominently as the experimental results themselves.

Solvent conditions

Folding rates are sensitive to solvent conditions, such as pH, buffer composition, and ionic strength, and thus a consensus set of solvent conditions would be helpful for purposes of comparing results obtained from different laboratories. Identifying a single set of conditions suitable for the characterization of all proteins, however, is not possible. For example, whereas pH 7.0 has been widely employed in the previous literature, is not suitable for some experiments (e.g., some hydrogen exchange experiments) and some proteins (e.g., those that are poorly soluble at neutral pH). Nevertheless, in an effort to reduce unnecessary complexity in interlaboratory comparisons and maximize backward compatibility with the literature, we recommend that pH 7.0 buffers be employed as a standard unless otherwise justified. It is similarly difficult to recommend a single buffer that is universally applicable to all protein folding experiments. We suggest, however, that 50 mM phosphate (although its pK_a is somewhat ionic strength dependent, and it binds some proteins, thereby altering their folding) (Chiti et al. 1998) or 50 mM HEPES is a reasonable choice that buffers well at neutral pH. Lastly, proteins are polyelectrolytes, and thus their stability and folding rates are a function of the solvent ionic strength (e.g., Went et al. 2004). Folding rates might ideally be determined under physiological ionic strength, but this is problematic as physiological conditions vary in different organisms. Additionally, high concentra-

tions of salts render some experiments, such as far-UV CD, difficult. We thus adopt the standard that, unless otherwise justified, no salt will be added beyond that provided by the buffer. For ease and accuracy in cross-group comparisons, the pH at which the experiments are conducted and the buffer system employed (including any additional salts) should be reported as clearly and prominently as the experimental results themselves.

Data reporting

The experimental folding community long ago adopted a de facto set of parameters to describe most protein folding reactions. Here we review these parameters and encourage their accurate and consistent reporting. We also argue that the pertinent structural characteristics of the experimental construct be clearly and completely described.

Folding phases exhibiting linear chevron diagrams

The estimated forward and backward rate constants for the kinetically-resolvable phases are extrapolated to estimate rate constants in the absence of denaturant. For phases that do not appear to exhibit kinetic rollover (phases for which the arms of the chevron plot are linear), a simple linear extrapolation, as described below, is appropriate. For experiments employing chemical denaturants and for which the arms of the chevron are linear, the so-called m -value (Pace and Shaw 2000) should be reported. This folding (or unfolding) m -value is the derivative of the natural logarithm of the folding (or unfolding) rate constant with respect to the denaturant concentration. Fitted values have historically been reported in units of kJ/mol/M for ease in comparison to equilibrium parameters. We recommend this practice be adopted as standard.

Folding phases exhibiting nonlinear chevron diagrams

Proteins that exhibit kinetic rollover are a significant complication. Rollover can arise due to the formation of stable on-pathway intermediates (e.g., Capaldi et al. 2002), the formation of stable off-pathway intermediates (e.g., Mogensen et al. 2004), a denaturant-induced shift between alternative pathways or change in the nature of the barrier (transition-state movement) (e.g., Otzen et al. 1999; Hedberg and Oliveberg 2004), ionic strength effects when guanidinium salts are used (e.g., Went et al. 2004; de los Rios and Plaxco 2005), or aggregation (e.g., Silow and Oliveberg 1997; Went et al. 2004). Clearly there is no universal fitting approach that is appropriate for all nonlinear chevrons. For this reason we suggest a fair default position for nonlinear chevrons would be to report both second-order polynomial extrapolations of the full data set (described below) and linear extrapolations of the linear regions close to the midpoint; both parameter sets should be clearly labeled as such.

The existence of multiple, mutually exclusive mechanisms for roll-over, however, suggests that it is also important to report the raw kinetic data [e.g., $\ln(k_{obs})$] as a function of denaturant concentration in numerical (tabular) form so that future researchers can refit the data as new models arise. We realize that the latter suggestion is effectively unprecedented in the literature, but as our understanding of the spectrum of folding mechanisms improves, the reporting of raw kinetic data represents an increasingly important contribution to the field. Such data can often be posted online as supplementary material, as is done here.

Non-two-state proteins

While progress has been made in terms of empirically or theoretically predicting two-state folding rates (such as via relationships between rates and native state topology) (e.g., Plaxco et al. 1998b), similar relationships between folding rates and equilibrium properties for non-two-state folding have generally not been found (for counter-examples, see the work of Ivankov et al. 2003, Kamagata et al. 2004). This no doubt at least partly results from the lack of a large, consistent, and complete data set of non-two-state folding kinetics. Moreover, since many different mechanisms can be used to describe such data, the physical nature of each of the phases in non-two-state folding is often unclear, further clouding any putative relationships between the equilibrium and kinetic properties of non-two-state proteins. Obviously there is no simple solution to this problem, but we believe that improvements in data reporting would provide an excellent opportunity to create a more useful data set for such studies. We thereby suggest that, for non-two-state proteins, investigators clearly and concisely identify the method via which the phase was observed (and the given rate obtained) and indicate, if possible, the proposed molecular origins of the phase and the range of conditions under which it was observed. In addition, the estimated (extrapolated) rate in the absence of denaturant and the denaturant m -value for each phase should be reported as described per two-state proteins.

Equilibrium parameters

A measure of whether the reported kinetic phases capture the entire folding and unfolding process is provided by a recapitulation of the folding free energy and overall energetic m -value as derived from a chevron plot with the values determined from equilibrium experiments. Thus, we consider it important that kinetic data are reported in conjunction with folding free energies and m -values derived via equilibrium unfolding experiments conducted under conditions identical to those employed in the kinetic study or from the end-point signals from kinetic amplitude analysis. The de facto community standard for fitting equilibrium unfolding curves is described in detail by Santoro and Bolen (1988) and repeated here for convenience. Spectroscopic

signals, f , (e.g., from fluorescence or CD) as a function of denaturant should be fitted to the following expression:

$$f = \frac{(a_1 + b_1[\text{denaturant}]) + (a_2 + b_2[\text{denaturant}]) \exp((\Delta G_u - m_{eq}[\text{denaturant}])/RT)}{1 + \exp((\Delta G_u - m_{eq}[\text{denaturant}])/RT)} \quad (1)$$

where ΔG_u is the free energy of unfolding; m_{eq} is the equilibrium m -value; a_1 and a_2 are the spectroscopic signals of the folded and unfolded states, respectively, in absence of denaturant; and b_1 and b_2 are the changes in the signals of the folded and unfolded states as function of denaturant concentration and account for sloped baselines.

Units

The rates of unimolecular folding reactions are conveniently reported in sec^{-1} . As described below, however, rate data are better fitted as the logarithm of observed rates. We thus suggest reporting natural logarithm (\log_e or \ln) of folding rate (in units of sec^{-1}) with the corresponding estimated errors in this parameter. With regard to protein stability and m -values, our consensus is that it is probably time that folding studies began to employ IUPAC (International Union of Pure and Applied Chemistry)-approved units appropriate for physical chemistry, namely, kJ/mol and kJ/mol/M , rather than kcal/mol and kcal/mol/M . Again, we believe that clearly indicating the units employed is far more critical than using any particular set of units. We note that the problem of ambiguously defined units (and the base of any logarithms employed) has historically been particularly acute in figure legends and labels.

Structural description

Given the importance of empirical and theoretical relationships between folding rates and structural characteristics such as length and topology, we believe that it is of the utmost importance to unambiguously describe the structure of the protein construct characterized. At a minimum, the full length (and, preferably, sequence) of the actual construct under investigation should be reported. This is most important in light of the fact that it has recently been shown that the addition of even small amino- or carboxyl-terminal tags can affect unfolding and refolding kinetics (Hamill et al. 1998; Went et al. 2004). It is also of value to denote the length of the "structured" region of this construct (i.e., the length excluding terminal tails that are "unstructured" in the native state). Lastly, it is of great utility to direct readers to the most relevant crystallographic or NMR-based structure file in the Protein Data Bank, as well as to indicate the appropriate chain designator and the residue numbers of the first and last structured residues in the file that are also in the characterized construct.

Data reduction

Estimated folding and unfolding rate constants in water, as well as m -values, are fitted parameters derived from experimental observations. The precise values of these fitted parameters depends on the fitting algorithms employed. After careful consideration, we recommend the following fitting procedures.

Determination of rates and m -values for linear chevrons via chemical denaturation

The most appropriate method of data analysis would be a simultaneous fitting of all of the observed signals (fluorescence, far-UV CD, etc.) as a function of time and denaturant concentration. However, this approach is so cumbersome that it has only rarely (Zitzewitz et al. 1995; Raschke et al. 1999) been employed. Instead, fitted relaxation rate constants (k_{obs}) are themselves typically fitted as a function of denaturant concentration (the so-called chevron plot, which is the plot of $\ln(k_{obs})$ versus denaturant concentration at a given temperature, pH, etc) to one of several well-established relationships. For chevrons with linear folding and unfolding arms, we recommend fitting $\ln(k_{obs})$ to the following equation:

$$\ln(k_{obs}) = \ln[\exp(A + m_f[\text{denaturant}]/RT) + \exp(B + m_u[\text{denaturant}]/RT)] \quad (2)$$

where m_f and m_u are the kinetic folding and unfolding m -values (the slopes of the two arms of the chevron, measured in kJ/mol/M) and $A = \ln(k_f^0)$ and $B = \ln(k_u^0)$, where k_f^0 and k_u^0 are the folding and unfolding rate constants, respectively, in the absence of denaturant. While fitting k_{obs} (the sum of two exponentials) is superficially equivalent to fitting $\ln(k_{obs})$ (the log of a sum), the two approaches are not mathematically equivalent. Because k_{obs} is an exponential function of denaturant concentration, parameters estimated by directly fitting k_{obs} are very heavily influenced by the most rapid observed rates. Because of technical limitations, however, errors in k_{obs} tend to be roughly proportional to k_{obs} , and thus the most rapid rates are usually the most poorly defined. By instead fitting $\ln(k_{obs})$ rather than k_p we put the experimental errors on a common scale and reduce the influence of the most poorly measured data on the parameter estimates. Moreover, since most current theoretical models predict $\ln(k_p)$ (i.e., barrier heights) rather than k_p , fitting to $A = \ln(k_f^0)$ and $B = \ln(k_u^0)$ provides direct estimates of the relevant parameters and their associated errors. For these reasons we believe it is more appropriate to fit $\ln(k_{obs})$ than to fit k_{obs} .

Determination of rate constants and m -values for chevrons with rollover

Some proteins exhibit significantly nonlinear behavior in the chevron diagram. A generic approach for fitting these

nonlinear chevrons is not obvious, since the extracted parameters will be model dependent. As described above, a simple, if perhaps unsatisfying, solution is to fit only the linear regions of the chevron close to the midpoint. We believe this should be done if a linear, two-state model is supported by the data over this range of denaturant concentrations; e.g., if $RT\ln(k_u/k_f)$ is within error of the equilibrium unfolding-derived ΔG_u . Parameters obtained from polynomial fittings of the entire data set

$$\ln(k_{obs}) = \ln[\exp(A + m_f[\text{denaturant}]/RT + m_f'[\text{denaturant}]^2) + \exp(B + m_u[\text{denaturant}]/RT + m_u'[\text{denaturant}]^2)] \quad (3)$$

can also be reported to describe the shape of the chevron plot. On the other hand, while equation 3 provides an empirical fit to the chevron data, it lacks a physical justification and thus probably should be employed with caution. More generally, we encourage the reporting of the raw $\ln(k_{obs})$ versus denaturant concentration values (published here as Supplemental Material) so that alternative mechanistic models of nonlinearity can be readily tested.

Error analysis

Standard errors, which quantify the uncertainty in parameter estimates, should also be reported. This will enable future researchers to appropriately weight folding observations by their estimated error when, for example, $\ln(k_p)$ is used as a dependent variable in an association study such as those relating folding rates to measures of protein topology. In addition, many other statistics can be derived from the raw data. Given that, under many circumstances, more detailed statistical analysis is required, we reiterate our strong recommendation that raw folding data be reported. Lastly, given that most theories estimate $\ln(k_p)$ rather than k_p , it is the estimate of the former parameter, and its standard error, that should be determined and reported.

Results

In an effort to support our recommendations with concrete action, we report here the kinetic and equilibrium folding parameters for 30 apparently two-state, single-domain proteins or protein domains (excised from larger proteins) characterized under the consensus conditions (Table 1). The folding kinetics of seven of these have not previously been reported. For many others, the folding kinetics have not been reported under the conditions recommended here. As per our recommendations, we also report the full structural details of the characterized constructs (Table 2), the full sequence of the characterized construct (Supplemental Material), and the data from which the kinetic parameters were derived (Supplemental Material). We report these data in

Table 1. Two-state data set

Protein	$\ln(k_f)$	$\ln(k_u)$	m_f kJ/mol/M	m_u kJ/mol/M	ΔG_u kJ/mol	m_{eq} kJ/mol/M	Conditions ^a
Abp1 SH3	2.46 ± 0.06	-2.72 ± 0.08	-6.36 ± 0.17	0.83 ± 0.04	13.0 ± 0.8	-6.9 ± 0.3	Phosphate/GuHCl/0.1 M NaCl
ACBP	6.96 ± 0.04	-3.86 ± 0.02	-9.74 ± 0.09	4.80 ± 0.09	23.7 ± 0.7	-13.1 ± 0.3	HEPES/GuHCl
ADAh2	6.80 ± 0.12	-0.42 ± 0.18	-3.12 ± 0.13	1.12 ± 0.05	17.1 ± 0.4	-3.9 ± 0.1	Phosphate/urea
Apo-azurin	4.91 ± 0.09	-4.02 ± 0.23	-8.03 ± 0.23	5.06 ± 0.19	29.2 ± 1.5	-17.6 ± 0.9	5 mM phosphate/GuHCl
CheW	7.44 ± 0.31	-12.05 ± 0.29	-8.92 ± 0.29	5.01 ± 0.16	50.2 ± 0.6	-14.6 ± 0.2	Phosphate/GuHCl
C12	5.75 ± 0.17	-10.33 ± 0.56	-5.70 ± 0.18	4.20 ± 0.25	32.5 ± 1.4	-8.3 ± 0.3	Phosphate/GuHCl
CTL9	3.27 ± 0.06	-7.85 ± 0.37	-3.09 ± 0.04	1.38 ± 0.10	27.2 ± 0.3	-4.4 ± 0.1	20 mM phosphate/pH 8/0.1 M NaCl/urea
EC298	9.08 ± 0.12	4.49 ± 0.61	-8.16 ± 0.27	5.95 ± 0.37	11.4 ± 1.6	-14.1 ± 1.9	Phosphate/GuHCl/T-jump
FKBP12	1.60 ± 0.09	-8.10 ± 0.29	-5.07 ± 0.14	2.20 ± 0.11	23.4 ± 0.9	-6.3 ± 0.2	Phosphate/urea
Fyn SH3	4.88 ± 0.17	-4.34 ± 0.37	-4.92 ± 0.21	1.68 ± 0.17	20.3 ± 1.4	-6.1 ± 0.4	Phosphate/GuHCl
GW1	3.98 ± 0.17	-1.66 ± 0.18	-10.12 ± 0.84	4.63 ± 0.18	15.5 ± 0.3	-13.9 ± 0.3	Phosphate/0.1 M NaCl/1 mM DTT
Im7*	7.20 ± 0.05	2.34 ± 0.04	-4.68 ± 0.09	0.53 ± 0.02	11.7 ± 0.6	-4.7 ± 0.2	Tris/urea
Im9*	7.33 ± 0.02	-1.87 ± 0.05	-4.53 ± 0.01	0.26 ± 0.02	20.9 ± 0.6	-4.4 ± 0.1	Tris/urea
λ -repressor	10.38 ± 0.28	3.21 ± 0.19	-7.34 ± 0.52	2.86 ± 0.16	21.2 ± 1.9	-10.1 ± 0.4	20 mM phosphate/pH 8/GuHCl
L23	2.02 ± 0.06	-3.88 ± 0.07	-3.46 ± 0.09	1.17 ± 0.02	11.3 ± 1.1	-3.3 ± 0.5	Phosphate/urea
mAcP	-1.58 ± 0.18	-9.00 ± 0.35	-1.27 ± 0.08	4.12 ± 0.10	20.4 ± 0.2	-5.3 ± 0.5	Tris/urea
NTL9	6.55 ± 0.02	0.08 ± 0.13	-1.84 ± 0.04	0.71 ± 0.04	17.3 ± 0.2	-2.6 ± 0.0	20 mM Tris/0.1 M NaCl/urea
Protein G	6.30 ± 0.08	-1.72 ± 0.17	-7.03 ± 0.15	1.38 ± 0.12	ND	ND	HEPES/pH 7.5/GuHCl
Protein L	4.10 ± 0.09	-3.25 ± 0.10	-6.38 ± 0.19	2.08 ± 0.05	19.9 ± 0.9	-8.1 ± 0.4	Phosphate/GuHCl
raf RBD	8.36 ± 0.12	-2.77 ± 0.65	-3.39 ± 0.09	1.03 ± 0.20	26 ± 3	-4.1 ± 0.5	Phosphate/urea/1 mM DTT
S6	6.07 ± 0.21	-8.28 ± 0.49	-7.04 ± 0.28	3.13 ± 0.22	34.7 ± 0.4	-10.0 ± 0.1	Phosphate/GuHCl
Sho1 SH3	2.11 ± 0.23	-2.49 ± 0.20	-6.60 ± 0.86	3.09 ± 0.13	9.4 ± 0.9	-13.1 ± 1.1	Phosphate/GuHCl/0.1 M NaCl
Spectrin SH3	1.05 ± 0.01	-4.83 ± 0.07	-2.27 ± 0.01	0.55 ± 0.02	13.9 ± 0.3	-2.7 ± 0.1	Phosphate/urea
SrcSH2	8.74 ± 0.21	-3.48 ± 0.29	-4.25 ± 0.12	0.89 ± 0.09	31.0 ± 0.4	-5.6 ± 0.1	20 mM imidazole/0.1 mM TCEP/urea
Src SH3	4.36 ± 0.07	-1.27 ± 0.13	-4.19 ± 0.16	1.70 ± 0.07	15.9 ± 2.5	-6.7 ± 0.8	Phosphate/GuHCl
Tm1083	6.85 ± 0.59	-5.26 ± 0.44	-5.79 ± 1.08	3.28 ± 0.19	38.2 ± 0.2	-9.8 ± 0.7	Phosphate/0.1 M NaCl/GuHCl
U1A	4.62 ± 0.05 ^b	-11.72 ± 0.58 ^b	-4.75 ± 0.15 ^c	4.13 ± 0.16 ^c	34.8 ± 1.5	8.6 ± 0.7	Phosphate/GuHCl
Ubiquitin	7.33 ± 0.06	-6.84 ± 0.34	-5.66 ± 0.08	3.43 ± 0.15	34.2 ± 0.4	-9.1 ± 0.1	20 mM acetate/pH 5/GuHCl
Urm1	2.58 ± 0.04	-3.30 ± 0.08	-2.72 ± 0.07	2.04 ± 0.03	13.0 ± 1.9	-4.2 ± 0.5	Phosphate/urea
VlsE	2.03 ± 0.24	-8.47 ± 0.25	-9.27 ± 0.60	5.24 ± 0.17	23.7 ± 0.7	-14.7 ± 0.4	5 mM phosphate/urea

* Folding/unfolding rate in water estimated from extrapolation to zero denaturant. Linear extrapolations unless otherwise noted (see footnotes c and d).

^a All experiments were conducted at 25°C; Other conditions were as follows *unless otherwise noted*: 50 mM buffer, no additional salts added. Lastly, all data were collected via stopped-flow dilution experiments unless otherwise noted.

^b Extrapolated rate in water estimated using a polynomial fit (Equation 3) due to significant curvature.

^c m -Values derived from the *linear* regions of a rather curved chevron plot.

tabular formats that we suggest efficiently convey all the key structural data, experimental conditions, and kinetic and equilibrium results that will be of interest to researchers in the folding field.

Discussion

Here we have suggested a set of experimental conditions that we propose the field should, so far as practical, employ as “default” conditions for quantitative biophysical studies of protein folding. We describe each parameter (with justification), preferred units, fitting algorithms, error estimation methods, and a minimal structural description that we believe should be reported when describing experimental folding results. We also describe here the folding of 30 apparently two-state proteins and isolated protein domains (Table 1), including their complete structural description (Table 2), the sequences of the characterized constructs, and the ex-

perimental data from which the kinetic parameters were derived (both as Supplemental Material). We hope that this work and the establishment of this like-minded group of experimental protein folding investigators (a “foldeomics consortium” spanning 17 laboratories in seven countries) will serve as a first step in an ongoing effort to compile a data set of the folding properties of a large number of reversibly folding proteins.

We should note that, because the conditions described here were specifically chosen to maximize backward compatibility with the literature, our recommendations generally only lead to relatively small changes in the reported folding rates of most of the previously characterized proteins. For example, 13 of the proteins characterized here were included in an earlier 24-protein data set that was used to illustrate the relationship between an empirical measure of topology termed relative contact order and two-state folding rates (Plaxco et al. 2000). Many of the rates reported here are effectively unchanged between this work and the previ-

Table 2. Structural information

Protein	PDB file ^a	Chain/ model ^b	Structure		Length of construct ^d
			Residue number ^c		
			Start	Stop	
ABP1SH3	1JO8		1	58	68
ACBP	1NT1		1	86	86
ADAh2	1O6X	1	1	81	81
Apo-azurin	1E65		1	128	128
CheW	1K0S	1	9	151	151
CI2	2CI2	1	20	83	64
CTL9	1DIV		58	149	92
EC0298	1JYG	—	1	69	89
FKBP	1FKF		1	107	110
FynSH3	1AVZ	C	85	141	78
GW1	1M9S		391	466	85
Im7*	1AYI		2	87	94
Im9*	1IMQ		2	86	93
λ-repressor	1LMB		6	85	80
L23	1N88		1	96	96
mAcP	1APS	—	1	98	99
NTL9	1DIV		1	56	56
Protein G	3GB1		1	56	62
Protein L	2PTL	1	18	77	79
raf RBD	1RFA	1	55	132	80
S6	1RIS		1	97	101
Sho1 SH3	NA		ND	ND	76
Spectrin SH3	1SHG		6	62	62
SrcSH2	1SPR	A	2	104	110
SrcSH3	1RLQ		9	64	61
Tm1083	NA		ND	ND	124
U1A	1URN		2	97	102
Ubiquitin	1UBQ		1	76	76
Urm1	NA		ND	ND	101
VlsE	1L8W	B	23	335	338

^a The Protein Data Base (PDB) file most relevant to the characterized construct.

^b Which chain in the PDB file (for files containing more than one polymer chain) is relevant or judged most representative.

^c PDB file residue numbers for the first and last "structured" (*i.e.*, showing a unique, well-defined conformation in the crystallographic or solution-phase structure) amino acids that are also present in the characterized construct.

^d The full length of the characterized construct, including unstructured amino- or carboxy-terminal tails (*e.g.*, uncleaved his-tags).

ous report, with the median rate variation being 24%, the mean variation 80%, and the largest variation a factor of six. Thus it is perhaps not surprising that the correlation between relative contact order and the $\ln(k_f)$ tabulated here, $r = 0.85$, is effectively indistinguishable from the $r = 0.86$ relationship derived using the originally reported rates (data not shown). Nevertheless, given the rather minimal effort required to adopt the recommended set of conditions, we believe that setting even (necessarily) arbitrary standards is appropriate even if the experimental variability the effort seeks to minimize is, itself, relatively small. More generally, we hope that adherence to a standardized, and less model-

dependent, system for measuring and reporting experimental results will stimulate and unify ongoing efforts to understand folding.

Materials and methods

The kinetic and thermodynamic characterizations were conducted under the conditions and by using the data reduction methods described above. Specific details regarding the production and characterization of the various proteins and isolated domains are described below, as are the initials of the specific investigators responsible for each individual data set.

SrcSH2 and SrcSH3

Both proteins were expressed with amino-terminal 6-his tags followed by a TEV protease site and were purified according to standard protocols (Qiagen). The tag was removed by overnight digestion with TEV protease at room temperature followed by a second passage over a Ni-NTA column. The SH2 domain was characterized under the conditions described (Table 1), which included 1 mM tris(2-carboxyethyl)phosphine (TCEP) (Pierce), to prevent the intermolecular disulfide bond formation. The SH3 domain was characterized under standard consensus conditions (Table 1). Kinetics of both proteins were determined by using a Biologic SFM 4 stopped-flow coupled to a Fluoromax 3 fluorimeter, with excitation at 290 nm and emission recorded between 310 and 330 nm. A slow, denaturant-independent refolding phase ($k_f = \sim 0.05 \text{ sec}^{-1}$) accounting for $\sim 15\%$ of the total amplitude of the SH2 domain folding was assumed to represent proline isomerization and ignored (data not shown). A slow phase ($\sim 0.02 \text{ sec}^{-1}$) observed in the folding of the srcSH3 domain and previously assigned (Grantcharova and Baker 1997) to proline isomerization was also ignored (D.W., S.M.).

NTL9 and CTL9

The proteins were expressed and purified as described (Sato et al. 2001) and characterization under the conditions reported (Table 1) as described (Kuhlman et al. 1998; Sato 2002; Sato and Raleigh 2002). CTL9 was studied at pH 8.0 because the stability and folding rate are strongly pH dependent between pH 5.0 and pH 7.5. The protein is most stable above pH 7.5 (Sato 2002; Sato and Raleigh 2002). The chevron plot represents the faster of two phases. Previous reports indicate that the slower phase reflects proline isomerism (Sato 2002; Sato and Raleigh 2002) (J.H., D.P.R.).

U1A and L23

U1A was expressed in *Escherichia coli* (C41) and purified on CM-Sepharose (Amersham Pharmacia) and S100 (Amersham Pharmacia). L23 was expressed in *E. coli* (BL21 DE3) and purified on SP-Sepharose (Amersham Pharmacia) and S100 (Amersham Pharmacia) and characterized as described (Hedberg and Oliveberg 2004). Stopped-flow measurements were carried out on an Applied Photophysics stopped-flow fluorimeter (SX17MV). Equilibrium fluorescence data were collected on J-810 JASCO spectrometer (excitation 280 nm, emission $>300 \text{ nm}$) (L.H., M.O.).

raf RBD

This protein was expressed, purified, and characterized as previously described (Vallée-Bélisle et al. 2004), save the characterization was performed under the conditions described (Table 1). The most rapid refolding phase, which accounts for ~40% of the total amplitude at low denaturant concentrations, was well fitted to a two-state folding transition. Three additional slower phases are resolved at below 3 M urea. These exhibit rates, amplitudes, and denaturant dependencies similar to those previously (Vallée-Bélisle et al. 2004) observed in GuHCl. In this previous study, rapid unfolding/refolding double-jump experiments suggested that the third and fourth transitions could be attributed to non-prolyl and proline residues isomerizations, respectively. However, the nature of the second transition remains to be conclusively determined, since its amplitude remains constant even when the delay between unfolding and refolding is insufficient to equilibrate the *cis-trans* isomers of the unfolded state (S.W.M., A.V.B.).

Urm1

The *Saccharomyces cerevisiae* protein was cloned from yeast genomic DNA into a prSET expression vector containing a 6-his affinity tag and a thrombin cleavage site. The protein was expressed in *E. coli* (C41), purified by Ni-NTA affinity column and thrombin cleaved at room temperature on the column. The cleaved protein was polished on a HiLoad 26/60 Superdex 75 (Amersham Pharmacia). Equilibrium data were acquired on an Aminco-Bowman Luminescence Spectrophotometer; final protein concentration was 1 μ M. Freshly thawed Urm1 was titrated into urea and 50 mM sodium phosphate (pH 7.0) and incubated for 4 h before measurement. Equilibrium and kinetic fluorescence measurements were monitored via tryptophan fluorescence by using an excitation wavelength of 280 nm. Kinetic data were acquired on an Applied Photophysics stopped-flow fluorimeter (SX-17MV); protein was at a final concentration of 1 μ M. Data were acquired above a wavelength of 320 nm by using a glass cut-off filter. Six traces were collected for each denaturant concentration, and data were first averaged and then fit to a single-exponential equation including a linear drift. Coincidence of equilibrium and kinetic data indicates that this protein folds via a two-state process (data not shown) (A.G.B., S.E.J.).

CheW

The protein, from *Thermotoga maritima*, was cloned into a modified pET15b vector and expressed with an amino-terminal 6-his tag followed by a TEV protease site and purified according to standard protocols (Qiagen). The tag was removed by overnight digestion with TEV protease at room temperature followed by a second passage over a Ni-NTA column. Unfolding and refolding rates were determined under standard consensus conditions by using an AVIV 202SF stopped-flow CD spectrometer. Equilibrium parameters were determined on an AVIV 62 CD spectrometer using a MicroLab titrator (D.W., S.M.).

ACBP

Bovine ACBP was obtained as previously described (Thomsen et al. 2002). Stopped-flow kinetic experiments were performed using a Biologic SFM-400 attached to a Jasco J-810. Excitation was at 280 nm, and fluorescence was measured with a 305-nm long-pass filter. The experiments were performed at 10 μ M protein concen-

tration under the conditions reported (Table 1). Equilibrium unfolding data were collected under the same conditions using a Perkin-Elmer LS-50b fluorimeter with excitation and detection at 280 nm and 356 nm, respectively (K.T., B.B.K., F.M.P.).

S6

A 100- μ M stock of *Thermus thermophilus* S6, obtained as previously described (Miller et al. 2002), was diluted into buffered GuHCl solutions under the conditions listed (Table 1) using a Biologic SFM-4, and fluorescence emission at 350 nm using 280 nm excitation was recorded using a Jobin-Yvon Horiba Fluoromax 3. Data collected below 2.02 M GuHCl were fit to a double exponential, and the major refolding phase was used to fit the chevron. Similar results have been observed for S6 previously (Otzen et al. 1999). All other points fit well to a single exponential equation (E.J.M., S.M.).

mAcP

The human protein was expressed and purified as described previously (Taddei et al. 1996). Kinetic unfolding/refolding was followed by using a Biologic SFM-3 stopped-flow fluorimeter, with excitation at 280 nm and a bandpass filter to monitor emitted fluorescence >320 nm under conditions listed (Table 1). Equilibrium urea denaturation was observed by monitoring the intrinsic fluorescence of 30 samples equilibrated in urea concentrations ranging from 0–8 M using a Perkin-Elmer LS 55 spectrofluorimeter with excitation wavelength at 280 nm and emission at 335 nm. Tris buffer was used rather than the recommended phosphate and HEPES because both bind to the native state of this protein with consequent alteration of its stability and kinetics (Chiti et al. 1998) (F.B., F.C.).

CI2

Stopped-flow kinetic folding and unfolding were performed with a Biologic SFM4 machine. Rate constants were determined by monitoring the change in the fluorescence signal. The excitation wavelength was 280 nm, and emission was collected at 320 nm. The faster measurable kinetic phase, which appears to represent a two-state, non-proline-limited phase (Jackson and Fersht 1991a), was used in this analysis. The equilibrium unfolding of the protein was characterized with an Aviv titrating CD spectrometer set at 220 nm. Equilibration time was 2 min. Solvent conditions were as described (Table 1) (N.-D.V., Y.B., M.A.R., K.W.P.).

Im7 and Im9**

These homologs were prepared, and their folding and unfolding kinetics and thermodynamics were measured as previously described (Capaldi et al. 2002; Friel et al. 2003) save under the conditions described here (Table 1). Each protein contains the amino-terminal tag ME(H)₆, which is denoted by the asterisk (i.e., Im7* and Im9*). Under the conditions employed, the folding of both proteins appear two-state (Fig. 1). A mild nonlinearity in the urea dependence of $\ln(k_f)$, which was previously noted (Ferguson et al. 1999) in the folding of Im7, was also observed under these consensus conditions (C.T.F., S.E.R.).

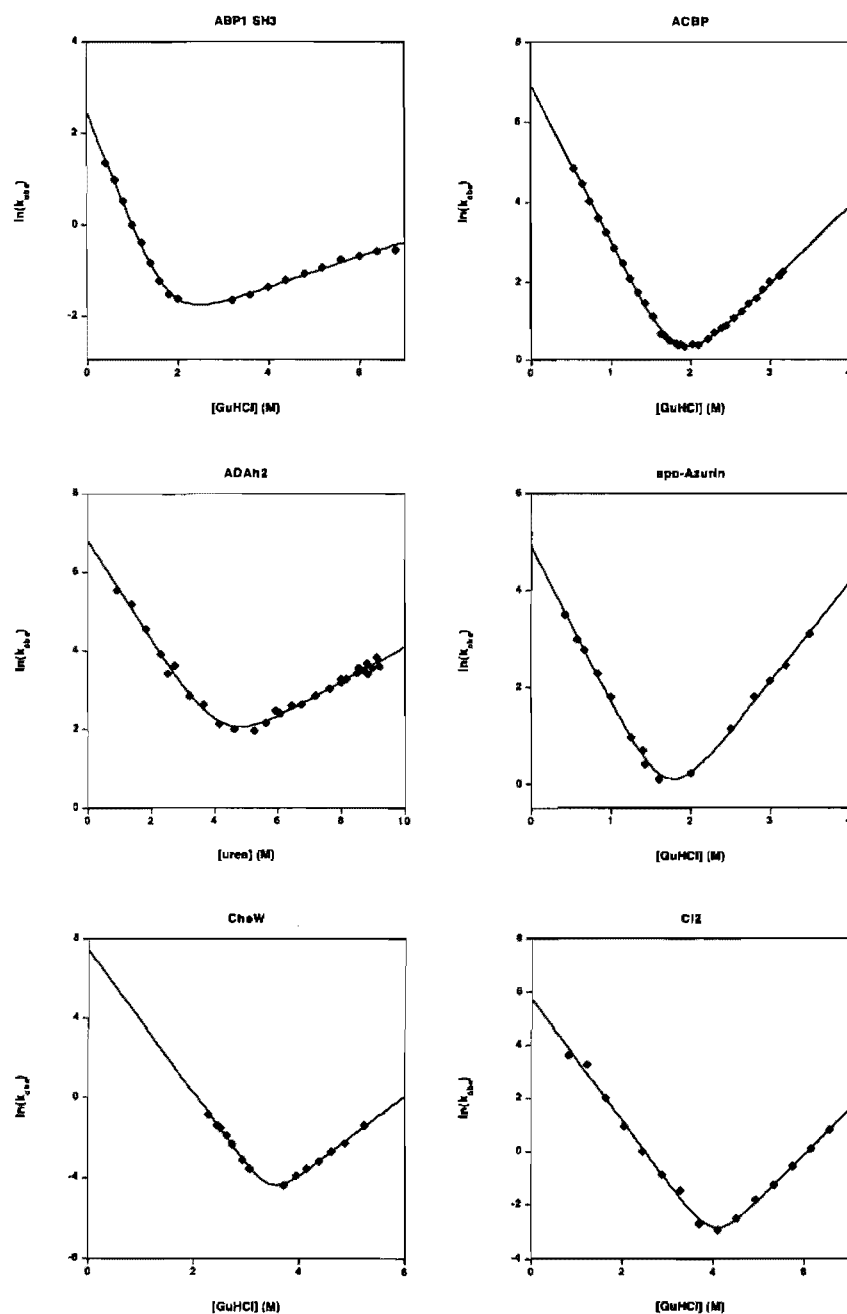


Figure 1. (Continued on next page)

EC0298

The protein was cloned into pet15B (which has a thrombin cleavage site) and expressed and purified as previously described (Maxwell et al. 2003). Folding/unfolding rates were determined under the conditions described (Table 1) by laser T-jump fluorescence spectroscopy (Qiu et al. 2002). Equilibrium free energy was obtained from a GuHCl titration as monitored by CD. As monitored

by the coincidence of the kinetically and equilibrium-derived folding free energy, this protein folds via an apparently two-state process (L.Q., D.B., K.L.M., S.J.H.).

FKBP12

This protein was purified and characterized as described previously (Main et al. 1999) save under the conditions described (Table 1) (E.R.G.M., S.E.J.).

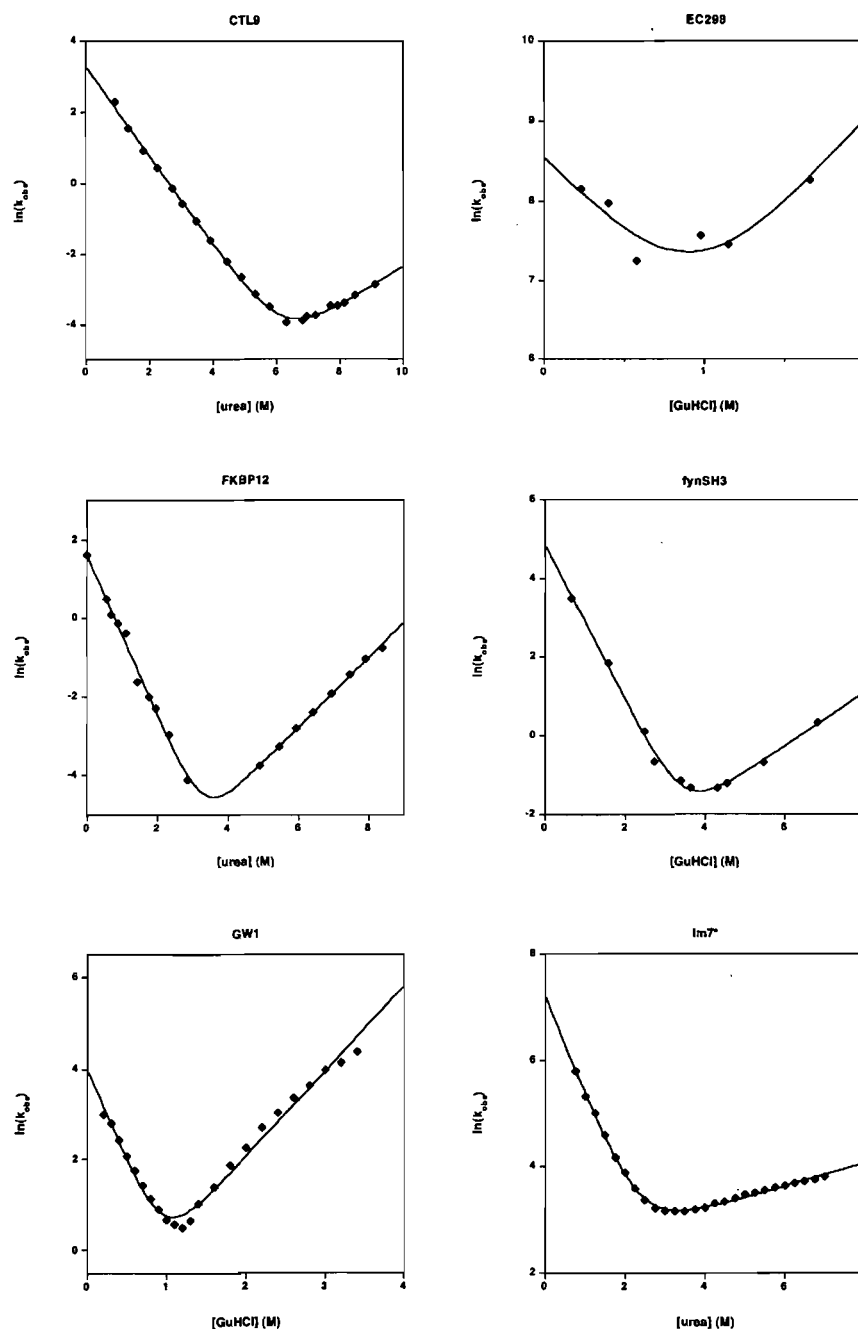


Figure 1. (Continued on next page)

Protein L and FynSH3

These proteins were expressed and purified as previously described (Scalley et al. 1997; Maxwell and Davidson 1998) and employed without cleavage of the his-tag. Kinetics were monitored via tryptophan fluorescence (excitation 280 nm, emission >320 nm) on an APP stopped-flow fluorimeter. Equilibrium unfolding was monitored via CD at 220 nm on an Aviv titrating CD spectrometer with a 2-min equilibration time. Buffer conditions were as

described (Table 1). Both proteins have previously been established to fold via an apparently two-state process (Scalley et al. 1997; Plaxco et al. 1998a) (M.A.R., K.W.P.).

Abp1 SH3, Sho1 SH3, GW1, and TM1083

The proteins were expressed in *E. coli* BL21* and purified under denaturing conditions (6M GuHCl) using nickel affinity chroma-

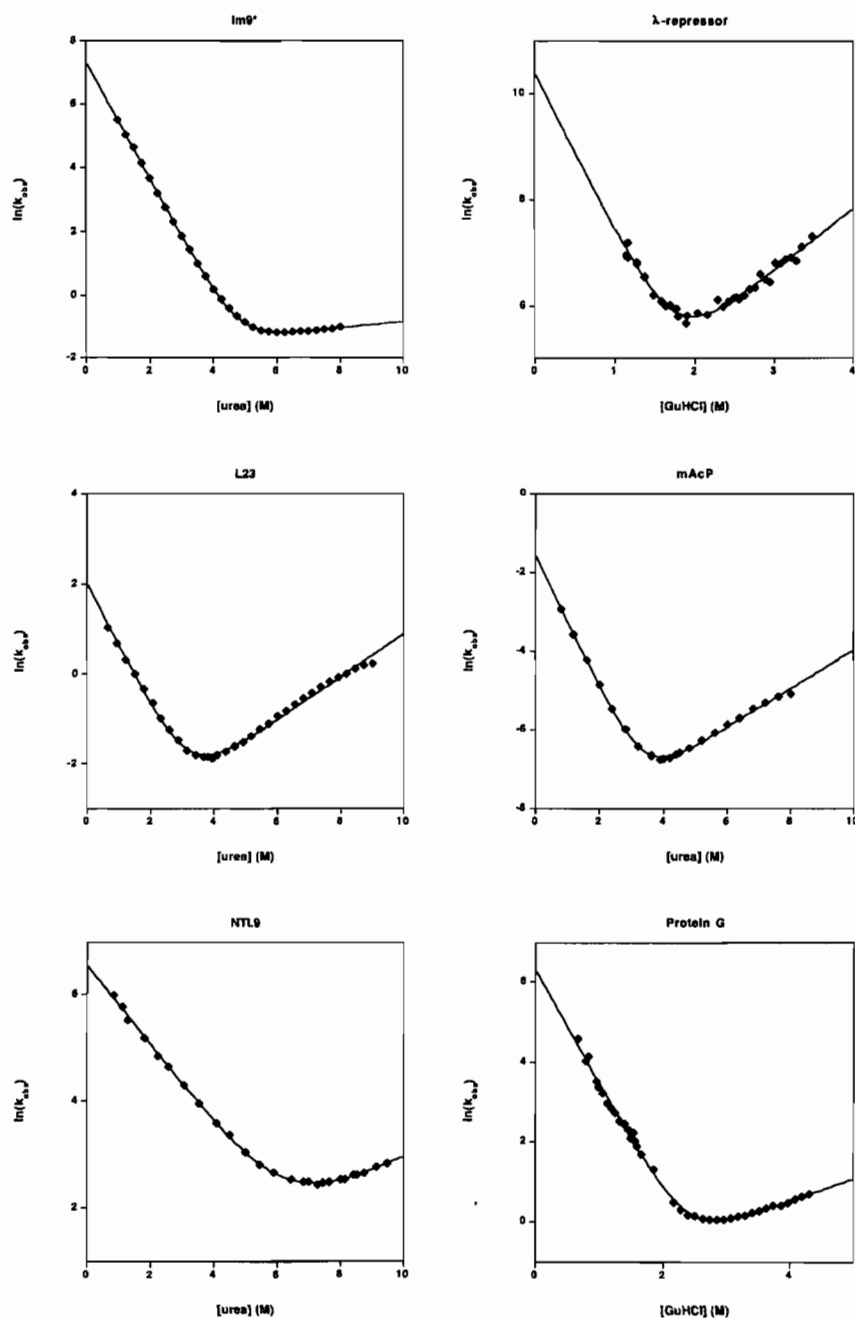


Figure 1. (Continued on next page)

tography by analogy to previous reports (Maxwell and Davidson 1998). The purified proteins were subsequently folded through dialysis in 50 mM phosphate buffer/100 mM NaCl and used as such without cleaving the tag. Folding and unfolding rates were obtained on a Biologic SFM-4 stopped-flow fluorimeter, with excitation at 295 nm and detection of total fluorescence >309 nm. Traces were fit to appropriate single exponentials. At each concentration of GuHCl, at least five separate shots of 3–5 μ M protein were averaged. Equilibrium melts were monitored using 1 μ M

protein on an Aviv Spectrofluorometer ATF 105 monitoring fluorescence emission at 340 nm with excitation at 295 nm (A.Z.A., A.R.D.).

λ Repressor and protein G

These proteins were expressed, purified, and characterized as previously described (Krantz et al. 2002) save under the conditions

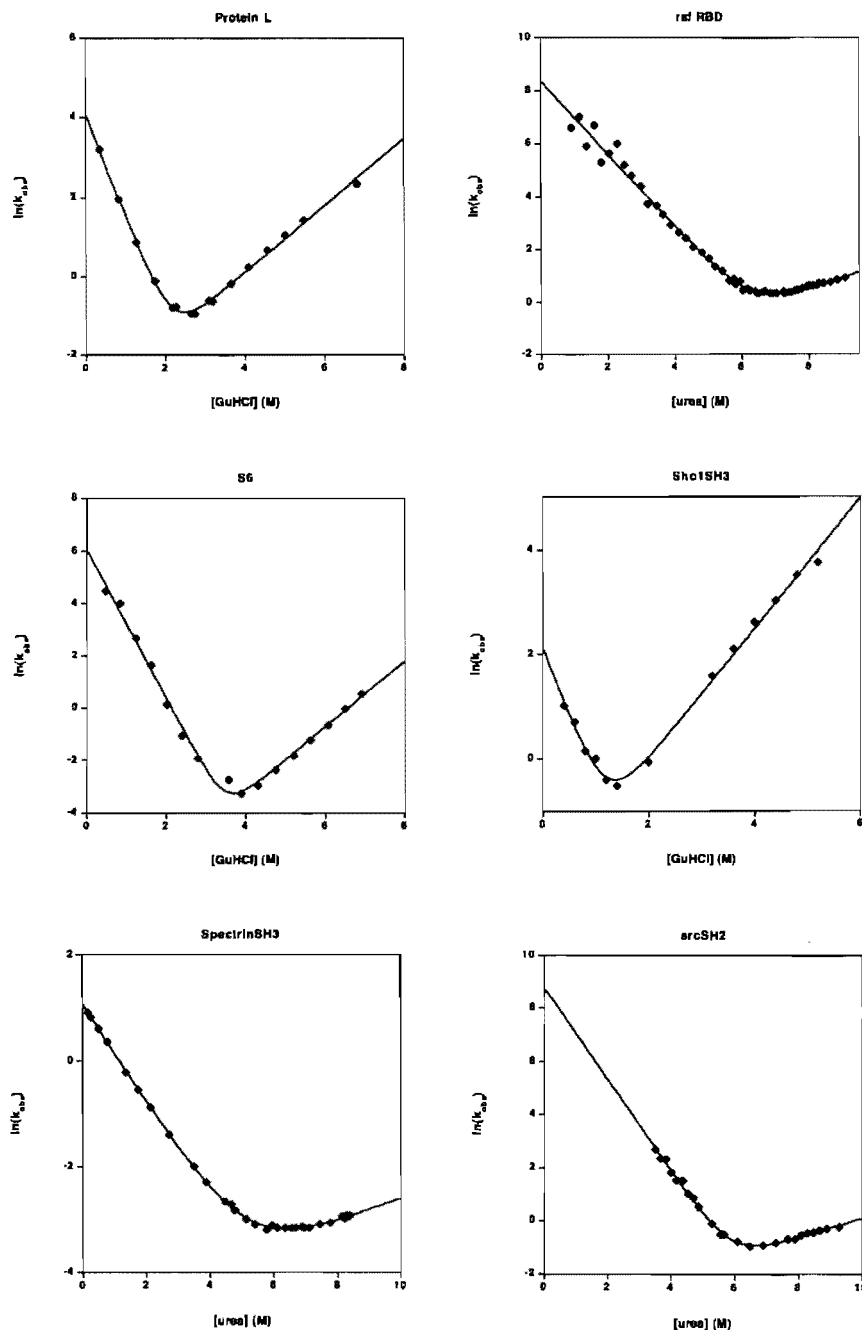


Figure 1. (Continued on next page)

described (Table 1). The pH employed for λ -repressor was raised slightly to improve its fluorescence properties (T.R.S.).

Ubiquitin

Data adopted from previously reported work (Krantz and Sosnick 2000) (T.R.S.).

SpectrinSH3 and ADAh2

Data for these proteins were adopted from previous work (Viguera et al. 1996; Villegas et al. 1998) and refitted as described here (L.S.).

Apo-azurin and VlsE

Apo-azurin is the metal (copper)-free version of native azurin. The data for both proteins were adopted from previous studies (Jones et

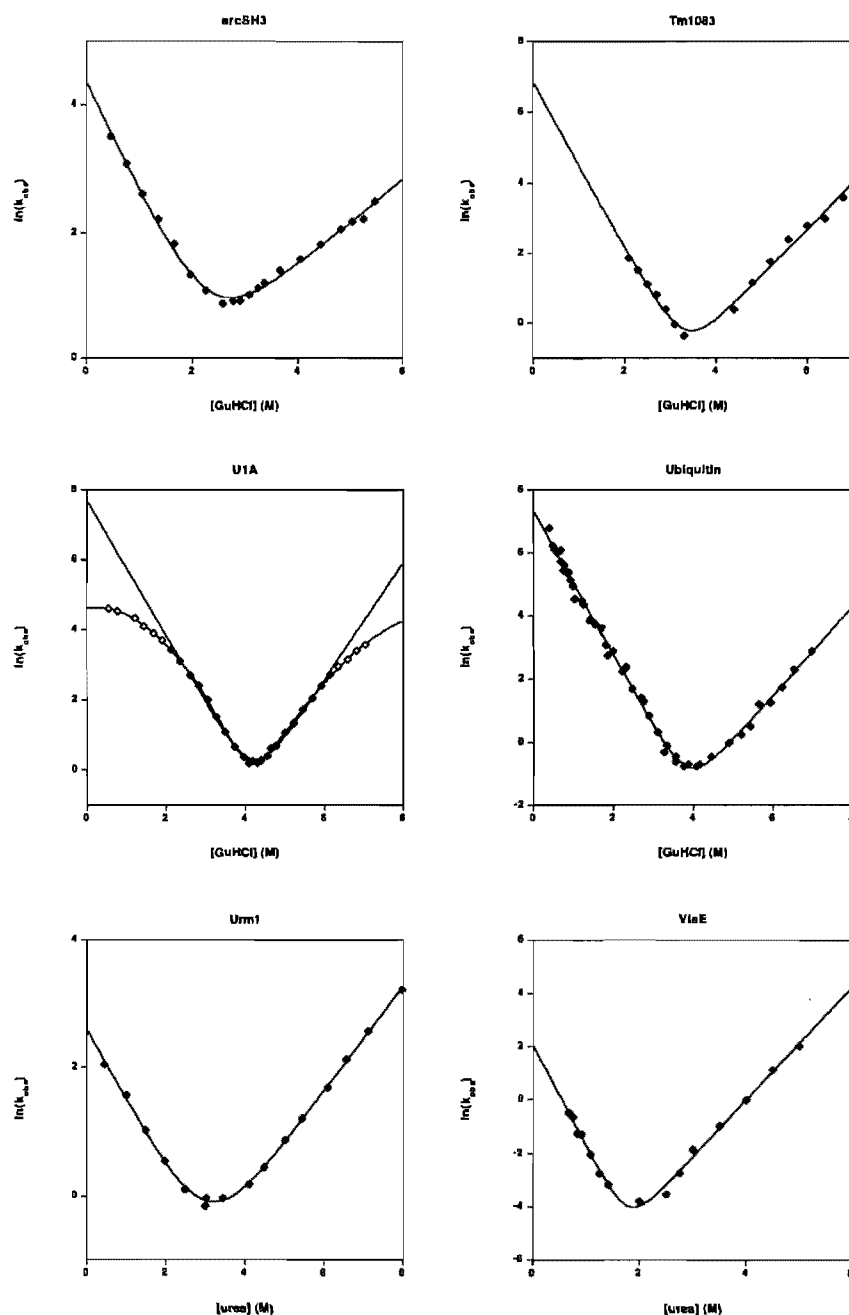


Figure 1. The refolding and unfolding kinetics of 30 small proteins and isolated protein domains. All but one fold with rates that are well-fitted ($r^2 > 0.97$) by linear-arm chevrons (equation 2) and appear to represent two-state folding under these conditions (as judged by the equivalence of the folding free energy as determined from kinetic and equilibrium chemical melts) (Table 1). The protein U1A exhibits significantly curved folding and unfolding arms under the conditions employed here (and other conditions)) (Otzen et al. 1999). Fitting the apparently linear region of the U1A chevron ($r^2 = 0.994$) produces an estimated $\ln(k_f)$ in water of 7.68 ± 0.18 ; fitting all of the data to a second order polynomial chevron ($r^2 = 0.9995$) produces the significantly lower estimate $\ln(k_f) = 4.62 \pm 0.05$. Several other proteins appear to exhibit curvature in their unfolding arms. But because these data are also well-fitted by equation 2, this curvature was ignored in the illustrated fits and in the data reported in Table 1. Note that the X- and Y-axis ranges vary significantly from plot to plot in this figure.

al. 2001; Pozdnyakova and Wittung-Stafshede 2001; Pozdnyakova et al. 2002; Jones and Wittung-Stafshede 2003) (P.W.S.).

Electronic supplemental material

We present as supplemental materials the full sequences of each of the constructs characterized in this study and the kinetic "chevron data" used to generate the reported kinetic parameters.

Acknowledgments

We acknowledge support from the following: the NIH via grants GM62868-01A2 (K.W.P.), GM59663 (P.W.-S.), GM50945 (S.M.), GM70941 (D.P.R.), and GM55694 (T.R.S.); the NSF via grant MCB-0347124 (S.J.H.); the Ontario Research and Development Challenge Fund (A.E.M.); the Canadian Institutes of Health Research (A.R.D.); and a faculty innovation award from the Johns Hopkins University (I.R.). S.E.R. and C.T.F. are members of the Astbury Centre for Structural Molecular Biology and also acknowledge the BBSRC and the Wellcome Trust for financial support. S.E.R. is a BBSRC Professorial Fellow. S.E.J. acknowledges support from the Welton Foundation and BBSRC. S.W.M. acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC) and is the Canada Research Chair in Integrative Genomics. E.R.G.M. holds a Tucker Prize Research Fellowship, Girton College, Cambridge. A.G.B. holds an EPSRC studentship. B.B.K., F.M.P., and K.T. acknowledge support from the John and Birthe Meyer Foundation and the Danish Research Council's Biotechnology Programme. K.L.M. is supported by a CIHR fellowship, and A.Z.A. is supported by an Ontario Graduate Student Scholarship.

Note added in proof

An updated version of the data presented here is also available online at <http://www.foldeomics.org>.

References

- Capaldi, A.P., Kleanthous, C., and Radford, S.E. 2002. Im7 folding mechanism: Misfolding on a path to the native state. *Nat. Struct. Biol.* **9**: 209–216.
- Chiti, F., Taddei, N., van Nuland, N.A., Magherini, F., Stefani, M., Ramponi, G., and Dobson, C.M. 1998. Structural characterization of the transition state for folding of muscle acylphosphatase. *J. Mol. Biol.* **283**: 893–903.
- de los Rios, M. A., and Plaxco, K.W. 2005. Apparent Debye-Huckel electrostatic effects in the folding of a simple, single domain protein. *Biochemistry* (in press).
- Ferguson, N., Capaldi, A.P., James, R., Kleanthous, C., and Radford, S.E. 1999. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **286**: 1597–1608.
- Friel, C.T., Capaldi, A.P., and Radford, S.E. 2003. Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **326**: 293–305.
- Grantcharova, V.P. and Baker, D. 1997. Folding dynamics of the src SH3 domain. *Biochemistry* **36**: 15685–15692.
- Hamill, S.J., Meekhof, A.E., and Clarke, J. 1998. The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin. *Biochemistry* **37**: 8071–8079.
- Hedberg, L. and Oliveberg, M. 2004. Scattered Hammond plots reveal second level of site-specific information in protein folding: $\phi^1 \beta^1$. *Proc. Natl. Acad. Sci.* **101**: 7606–7611.
- Ivankov, D.N., Alm, E., Plaxco, K.W., Baker, D., and Finkelstein, A.V. 2003. Contact order revisited: The influence of protein size on folding rates. *Protein Sci.* **12**: 2057–2062.
- Jackson, S.E. 1998. How do small single-domain proteins fold? *Fold. Des.* **3**: R81–R91.
- Jackson, S.E. and Fersht, A.R. 1991a. Folding of chymotrypsin inhibitor-2. 1: Evidence for a two-state transition. *Biochemistry* **30**: 10428–10435.
- . 1991b. Folding of chymotrypsin inhibitor-2. 2: Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition-state of folding. *Biochemistry* **30**: 10436–10443.
- Jones, K. and Wittung-Stafshede, P. 2003. The largest protein observed to fold by two-state kinetic mechanism does not obey contact-order correlation. *J. Am. Chem. Soc.* **125**: 9606–9607.
- Jones, K., Guidry, J., and Wittung-Stafshede, P. 2001. Characterization of surface antigen from Lyme disease spirochete *Borrelia burgdorferi*. *Biochem. Biophys. Res. Commun.* **289**: 389–394.
- Kamagata, K., Arai, M., and Kuwajima, K. 2004. Unification of the folding mechanisms of non-two-state and two-state proteins. *J. Mol. Biol.* **339**: 951–965.
- Krantz, B.A. and Sosnick, T.R. 2000. Distinguishing between two-state and three-state models for ubiquitin folding. *Biochemistry* **39**: 11696–11701.
- Krantz, B.A., Srivastava, A.K., Nauli, S., Baker, D., Sauer, R.T., and Sosnick, T.R. 2002. Understanding protein hydrogen bond formation with kinetic H/D amide isotope effects. *Nat. Struct. Biol.* **9**: 458–463.
- Kuhlman, B., Luisi, D.L., Evans, P.A., and Raleigh, D.P. 1998. Global analysis of the effects of temperature and denaturant on the folding and unfolding kinetics of the N-terminal domain of the ribosomal protein L9. *J. Mol. Biol.* **284**: 1661–1670.
- Main, E.R., Fulton, K.F., and Jackson, S.E. 1999. Folding pathway of FKBP12 and characterisation of the transition state. *J. Mol. Biol.* **291**: 429–444.
- Makhatadze, G.I. 1999. Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J. Phys. Chem. B* **103**: 4781–4785.
- Maxwell, K.L. and Davidson, A.R. 1998. Mutagenesis of a buried polar interaction in an SH3 domain: Sequence conservation provides best prediction of stability effects. *Biochemistry* **37**: 16172–16182.
- Maxwell, K.L., Bona, D., Liu, C., Arrowsmith, C.H., and Edwards, A.M. 2003. Refolding out of guanidine hydrochloride is an effective approach for high-throughput structural studies of small proteins. *Protein Sci.* **12**: 2073–2080.
- Miller, E.J., Fischer, K.F., and Marqusee, S. 2002. Experimental evaluation of topological parameters determining protein-folding rates. *Proc. Natl. Acad. Sci.* **99**: 10359–10363.
- Mogensen, J.E., Ipsen, H., Holm, J., and Otzen, D.E. 2004. Elimination of a misfolded folding intermediate by a single point mutation. *Biochemistry* **43**: 3357–3367.
- Otzen, D.E., Kristensen, O., Proctor, M., and Oliveberg, M. 1999. Structural changes in the transition state of protein folding: Alternative interpretations of curved chevron plots. *Biochemistry* **38**: 6499–6511.
- Pace, C.N. and Shaw, K.L. 2000. Linear extrapolation method of analyzing solvent denaturation curves. *Proteins* **4** (Suppl.): 1–7.
- Plaxco, K.W., Gujjarro, J.I., Morton, C.J., Pitkeathly, M., Campbell, I.D., and Dobson, C.M. 1998a. The folding kinetics and thermodynamics of the FynSH3 domain. *Biochemistry* **37**: 2529–2537.
- Plaxco, K.W., Simons, K.T., and Baker, D. 1998b. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**: 985–994.
- Plaxco, K.W., Simons, K.T., Ruczinski, I., and Baker, D. 2000. Sequence, stability, topology and length: The determinants of two-state protein folding kinetics. *Biochemistry* **39**: 11177–11183.
- Pozdnyakova, I. and Wittung-Stafshede, P. 2001. Copper binding before polypeptide folding speeds up formation of active holo: *Pseudomonas aeruginosa* azurin. *Biochemistry* **40**: 13728–13733.
- Pozdnyakova, I., Guidry, J., and Wittung-Stafshede, P. 2002. Studies of *Pseudomonas aeruginosa* azurin mutants: Cavities in β -barrel do not affect refolding speed. *Biophys. J.* **82**: 2645–2651.
- Qiu, L.L., Pabit, S.A., Roitberg, A.E., and Hagen, S.J. 2002. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μ s. *J. Am. Chem. Soc.* **124**: 12952–12953.
- Raschke, T.M., Kho, J., and Marqusee, S. 1999. Confirmation of the hierarchical folding of RNase H: A protein engineering study. *Nat. Struct. Biol.* **6**: 825–831.
- Santoro, M.M. and Bolen, D.W. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1: Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry* **27**: 8063–8068.
- Sato, S. 2002. "Folding of ribosomal protein L9 and its isolated N- and C-terminal domains." Ph.D. thesis, Department of Chemistry, State University of New York at Stony Brook, Stony Brook, NY.
- Sato, S. and Raleigh, D.P. 2002. pH-dependent stability and folding kinetics of a protein with an unusual α - β topology: The C-terminal domain of the ribosomal protein L9. *J. Mol. Biol.* **318**: 571–582.
- Sato, S., Xiang, S., and Raleigh, D.P. 2001. On the relationship between protein stability and folding kinetics: A comparative study of the N-terminal domains of RNase HI, *E. coli* and *Bacillus stearothermophilus* L9. *J. Mol. Biol.* **312**: 569–577.
- Scalley, M.L., Yi, Q., Gu, H.D., McCormack, A., Yates, J.R., and Baker, D.

1997. Kinetics of folding of the IgG binding domain of peptostreptococcal protein L. *Biochemistry* **36**: 3373–3382.
- Silow, M. and Oliveberg, M. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci.* **94**: 6084–6086.
- Spector, S. and Raleigh, D.P. 1999. Submillisecond folding of the peripheral subunit-binding domain. *J. Mol. Biol.* **293**: 763–768.
- Taddei, N., Stefani, M., Magherini, F., Chiti, F., Modesti, A., Raugei, G., and Ramponi, G. 1996. Looking for residues involved in the muscle acylphosphatase catalytic mechanism and structural stabilization: Role of Asn 41: Thr 42 and Thr 46. *Biochemistry* **35**: 7077–7083.
- Thomsen, J.K., Kragelund, B.B., Teilum, K., Knudsen, J., and Poulsen, F.M. 2002. Transient intermediary states with high and low folding probabilities in the apparent two-state folding equilibrium of ACBP at low pH. *J. Mol. Biol.* **318**: 805–814.
- Vallée-Bélisle, A., Turcotte, J.-F., and Michnick, S.W. 2004. raf RBD and ubiquitin proteins share similar folds, folding rates and mechanisms despite having unrelated amino acid sequences. *Biochemistry* **43**: 8447–8458.
- Viguera, A.R., Wilmanns, M., and Serrano, L. 1996. Different folding transition states could result in the same native structure. *Nat. Struct. Biol.* **3**: 874–880.
- Villegas, V., Martinez, J.C., Aviles, F.X., and Serrano, L. 1998. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.* **283**: 1027–1036.
- Went, H.M., Benitez-Cardoza, C.B., and Jackson, S.E. 2004. Is there an intermediate populated on the folding pathway of ubiquitin? *FEBS Lett.* **567**: 333–338.
- Zitewicz, J.A., Bilsel, O., Luo, J.B., Jones, B.E., and Matthews, C.R. 1995. Probing the folding mechanisms of a leucine-zipper peptide by stopped-flow circular-dichroism spectroscopy. *Biochemistry* **34**: 12812–12819.